

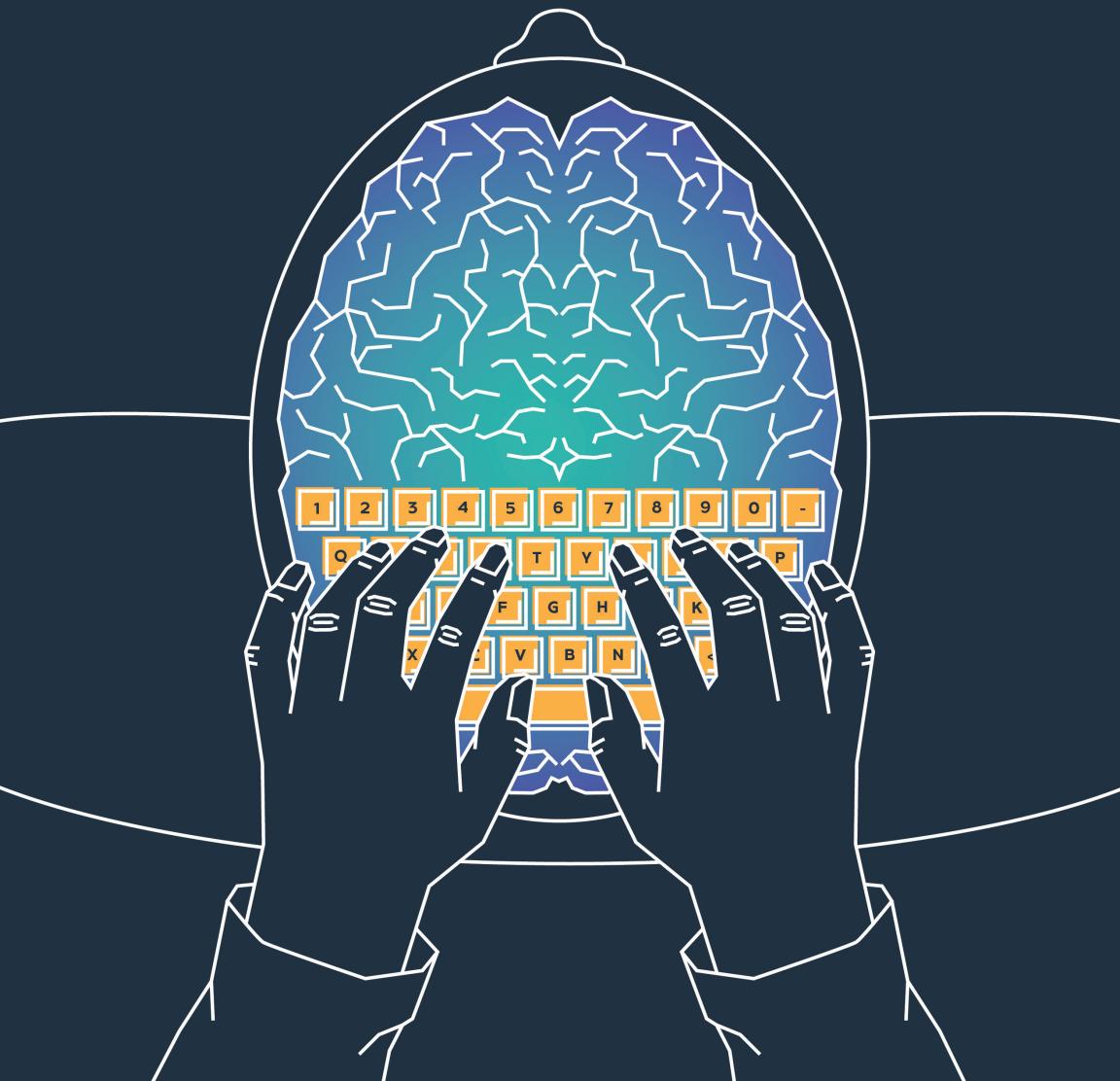
DEEP THOUGHT

ideas⁴²

||||||| A CYBERSECURITY STORY |||||

Alex Blau, Alexandra Alhadoff, Michael Stern, Scott Stinson, Josh Wright

ideas42.org/cyber



// ABOUT IDEAS42



IDEAS42 USES INSIGHTS FROM BEHAVIORAL SCIENCE to help solve difficult social problems and make an impact at scale. We grew out of research programs in psychology and economics at top academic institutions and our work draws on decades of experimental scientific research in decision making, along with the most rigorous methods in program and policy evaluation. We work in a number of areas, including consumer finance, economic mobility and opportunity, health, education, energy efficiency, and international development. The consequences of the behavioral issues we tackle are often profound. A failure to adhere to medication can be life-threatening. Dropping out of school can prevent a person from achieving her potential. All too often, the reasons for these failures turn out to be small and remediable—but also usually overlooked or dismissed as unimportant. Our approach involves carefully diagnosing the behavioral issues that prevent otherwise well-designed programs and products from achieving their goals. We identify subtle but important contextual details that can influence behavior, and design innovative solutions that help to overcome or amplify their effects. Our work involves a lot of observation, plenty of patience, and a willingness to be surprised. Most of all, though, it involves asking the right questions.

// ACKNOWLEDGMENTS

THIS PROJECT WOULD NOT HAVE BEEN POSSIBLE without the support of many individuals and institutions. ideas42 would like to thank the William and Flora Hewlett Foundation, whose generous support enabled us to examine critical challenges in cybersecurity through the lens of behavioral science. We would like to give particular thanks to our Program Officer, Eli Sugarman, who lent us his sagely advice and guidance throughout the project and helped us get up to speed quickly on this incredibly diverse topic area by connecting us with the right people and content. Additionally, many thanks to the Hewlett Foundation President, Larry Kramer, for having the foresight to push forward cybersecurity as a major foundation initiative so that it gets the attention it deserves as a critical issue for our country and humankind. We would also like to thank our colleagues at the New America Foundation including Ian Wallace, co-director of the cybersecurity program, and Robert Morgus, policy analyst, who provided us with their expertise throughout the project and connected us with many exceptional people. We would like to give special thanks to Ian in particular, who first suggested that we use the behavioral perspective to take a look at cybersecurity challenges—without that nudge, we may not have endeavored down this path in the first place. Many thanks to those who supported our writing effort including the narrative wisdom of Peter Singer and Peter Kelly as well as Jason Hong, Nicole Becher, and Greg Michaelidis who provided feedback on our early draft. We are also indebted to the 60 experts from academia, government, policy, consulting, and industry who graciously contributed their time and invaluable insights to this project, and whose experiences informed the narrative we've crafted. There are many more people who helped to contribute to this work whom we may not have named here and we thank them as well for all their support. Finally, we'd like to acknowledge the coders, end users, IT administrators, policy makers, and executives working at the forefront of cybersecurity who inspired this work.

Ideas42 © 2017 | ideas42.org

80 Broad Street, 30th Floor

New York, NY 10004



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](#).

// CONTENTS

CHAPTER 1: THE HACK.....	1
BOX: Open Wifi and Channel Factors.....	3
CHAPTER 2: PEOPLE, NOT COMPUTERS.....	9
BOX: A Brief History of Behavioral Economics.....	14
CHAPTER 3: TAKE ME OUT TO THE BALL GAME	17
BOX: The Affect Heuristic and Bad Links.....	19
BOX: Warnings, Habituation, and Generalizability.....	22
BOX: The Choice Architecture of Updates.....	24
BOX: Thinking Fast and The Internet.....	26
CHAPTER 4: D33PTHOUGH1	31
BOX: Incomplete Rules of Thumb	33
BOX: Status Quo Bias and Passwords	36
BOX: The Hassle with Multifactor Authentication.....	38
BOX: Availability and the Risks We See.....	40
CHAPTER 5: OPEN ACCESS.....	41
BOX: Vulnerability Sharing and Present Bias.....	44
BOX: Access Controls and The Context of Scarcity.....	47
BOX: Congruence Bias and Investment.....	51
CHAPTER 6: GONE PHISHING.....	55
BOX: Phishing from Authority.....	59
BOX: Primed to See What They Want You to See.....	61
BOX: Insecurity by Default.....	63

CHAPTER 7: THE WAGER.....**67**

BOX: Mental Models and Thinking About Security..... 72

CHAPTER 8: THE LONG WAY HOME.....**75****APPENDIX**

Updating.....	84
Security Warnings.....	87
Safe Coding.....	90
Passwords.....	95
Multifactor Authentication Use.....	100
Employee Commitment to Cybersecurity.....	103
Access Control Management.....	106
Threat and Vulnerability Sharing	110
End User Security Settings.....	114
Phishing.....	119
Investing in Cybersecurity.....	123

DEEP THOUGHT

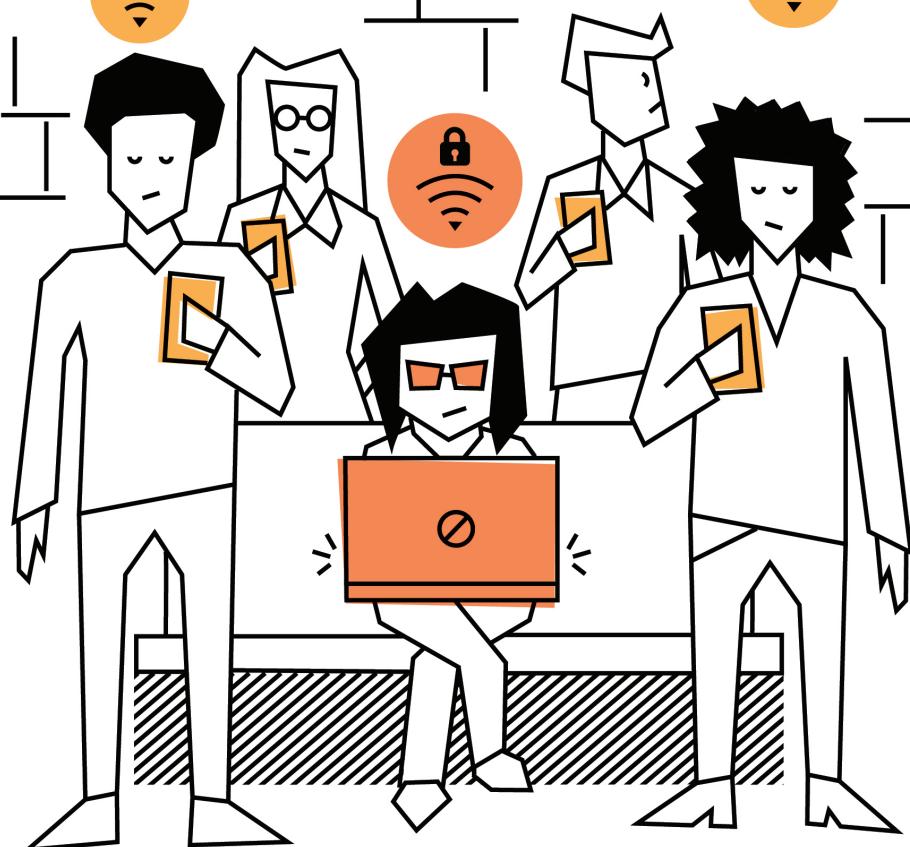
||||||| A CYBERSECURITY STORY |||||

DOWNTOWN
& BROOKLYN

N

Q

R



CHAPTER 1

THE HACK

THE RUMBLE OF THE APPROACHING N TRAIN echoed through the subway tunnels. Commuters, standing at the platform's edge, began leaning out to glimpse the train's light peeking out from around the bend. Others stood idly playing with their phones, reading books, or thumbing through papers in advance of the workday. But one woman, her face illuminated in cobalt, sat on a bench, hunched over her laptop screen.

The train arrived on the platform like a riot. The doors opened, and a mass of bodies exchanged places. Those who got off clamored towards the stairs leading up to the street, while those onboard pressed themselves into some uninhabited nook of humanity and held on. The doors of the train closed following a loud “ding,” and the train lurched back into motion, continuing its journey beneath the city.

However, the woman on the bench did not move. A dozen trains had come and gone while she sat there, but she paid as much attention to the bustle as a beach-goer would to the waves and rising tide. Instead, she continued to clack away on her keyboard, engrossed in her work, ignorant to the goings-on around her.

As the train moved out of earshot, the woman stopped to scrutinize the contents of the screen in front of her. Two hundred and eighty-seven words, honed like a knife, filled the page of her word processor. She silently mouthed the words to herself, weighing each sentence and tonguing each letter’s subtle edge. The cursor

blinked impatiently, awaiting its next command, but she was satisfied. There was nothing left to write, so she exported the document as a PDF to her desktop. The computer's clock read 8:51 AM; she had managed to buy herself a few moments of reprieve. Relaxing her focus, she noticed a growling in her stomach and the onset of a caffeine-deprived headache.

"Soon," she muttered to herself, coaxing her stomach silent. Self-care was often the sacrificial lamb when deadlines were tight, and today ended up being no exception to the rule.

The platform twice more filled and drained of commuters. Not one of them paid the woman on the bench any mind as she worked on what appeared to be a mundane, pre-workday task. At one point a man sat down next to her and pulled a computer out of his bag to do a bit of work between train rides. He was off again on another train as quickly as he had arrived, without as much as a glance in her direction. It wasn't uncommon to see people using the 5th Avenue 59th Street N station's notoriously reliable but often congested public Wi-Fi to squeeze a bit of work in during their commute. While a mobbed Wi-Fi network might be problematic for the average user, the woman sitting on the bench selected this station, and Wi-Fi network in particular because it afforded her the anonymity she needed. She glanced at the clock on the computer. Six minutes past nine. It was time to begin.

She went to work like a machine, toggling her virtual private network (VPN),ⁱ opening her Tor browser, turning on Wi-Fi, and connecting to the public network incognito. Rattling at her keyboard once again, she navigated to the target web page, logged in with stolen account credentials, and uploaded the document to the cloud. She paused, considering the risk one last time. She wondered what repercussions would come, and whether she would be able to keep herself out of the fray. Putting those fears aside, the muscles in her hand fired, sending the document out into the nexus. The first domino had fallen.

Another train pulled into the station, and this time, as the mob poured out of the car doors, she was subsumed by the crowd which flowed up the stairs like a torrent to the street.

ⁱ A virtual private network (VPN for short) allows users to send and receive data across the Internet securely. In other words, they simulate the function of connecting users directly to private networks. Among other benefits, VPNs protect identity and allow users to safely and remotely access company networks.



OPEN WI-FI AND CHANNEL FACTORS

Research from social psychology suggests that people have a tendency to take a given action when environmental factors either eliminate constraints or guide behavior toward that action.¹ Simply put, people do things when they are easy to do. Social psychologists call these catalyzing environmental factors **CHANNEL FACTORS** because they have a tendency to ‘channel’ people’s actions.²

Commonly-used public Wi-Fi networks, like the one accessed by the woman on the bench, represent a potentially dangerous type of channel factor. Open Wi-Fi networks channel people into connecting to them because they require no authentication to join. Some users may only be needed to select the Wi-Fi network to join, while others may connect automatically because of the existing security settings on their devices. However, while this ‘channel’ is convenient for the average computer user, it also presents an opportunity for hackers.³

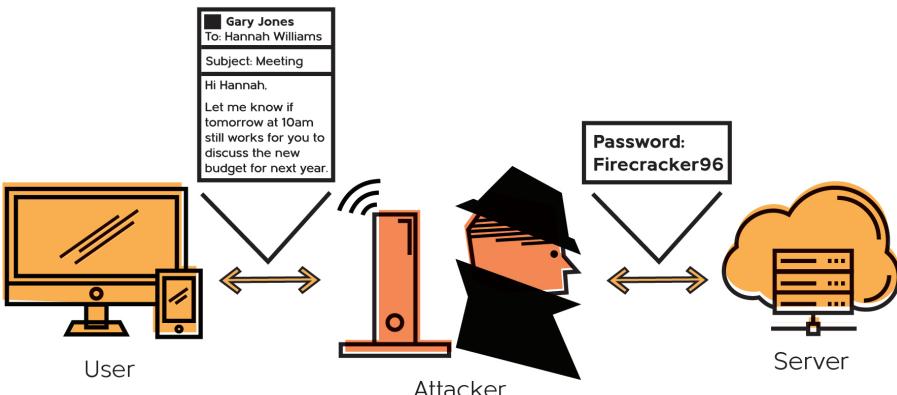
The real risk of open Wi-Fi networks is that hackers can position themselves between the network and users and execute a man-in-the-middle (MITM) attack, redirecting all network traffic through their computer before it goes on to the router. Doing so allows hackers to see all the information heading out into the Internet, including communications, personal information, and web traffic.

Device manufacturers hoping to reduce the likelihood that users will put themselves at risk should make connecting to open networks a little more hassle-filled. By developing more gating mechanisms such as forcing the user to acknowledge the security risks before joining, or turning off the functionality that allows users to connect automatically, it may be possible to nudge users away from using insecure networks.

Regardless, the next time you choose to connect to a public Wi-Fi network, remember that if it’s easy for you, it’s easy for the bad guy.



MAN IN THE MIDDLE ATTACK



IT WAS 10:27 AM and Damien’s office phone was ringing off the hook. News outlets from around the country had started calling him just after 10:00 AM to confirm the validity of a press release CapitalCorp had supposedly crossed over the wire, which was news to Damien. But the journalists, whose calls he stopped picking up altogether around 10:15, all said the release came from his CapitalCorp communications team. From the TV in his office, he gleaned, to his horror, that news outlets were already reporting on a lie that it was now Damien’s job to squash. A reporter, who appeared to be somewhere in the Stock Exchange building, was saying, “CapitalCorp’s CEO is stepping down following an admission of multiple instances of sexual misconduct in the office, the misuse of company funds to pay off the victims, and an associated accounting cover-up to hide his offenses.” Multiple news outlets were citing statements from his team, but no one on his staff seemed to know where those comments originated.

Damien’s cell phone vibrated in his jacket pocket. CapitalCorp’s CEO, James Robinson, was calling him directly.

“What the hell is going on?” James spat.

“I have no idea,” Damien said. “They’re saying we issued a press release, but we never—”

“You need to figure it out and get out in front of this!”

Damien ran the names of the communications staff through his head, wondering who in his department could have done something like this. “James, is there any truth here? You know if there is, you need to tell me so I can do my job.”

“Whose side are you on? Of course, none of it is true!”

“James, we’ll figure this out. It’s obviously a mistake.”

“Obviously?” James was incredulous. He groaned. “My wife’s calling. Damien, I have a thousand other calls I need to make right now. You need to get rid of this—yesterday!” The line went dead.

Damien slowly tilted his head into his hands. On the television, a market analyst was yelling “sell now,” his backdrop a cartoonish line graph of CapitalCorp’s plummeting stock price. Damien racked his brain for where the press release came from, how it got sent out without crossing his desk, and who was talking to reporters. He shook his head and picked up his phone to call Dana, CapitalCorp’s Chief Information Security Officer (CISO), about his audit request.

“We couldn’t find any outgoing emails that looked fishy, but it does look like your department’s VoIP lines weren’t accepting any inbound calls from around 9:15 to 10:00 AM,” Dana told Damien. “It doesn’t feel like a coincidence. In any event, I’ll keep the team on this and let you know as soon as we find something.”

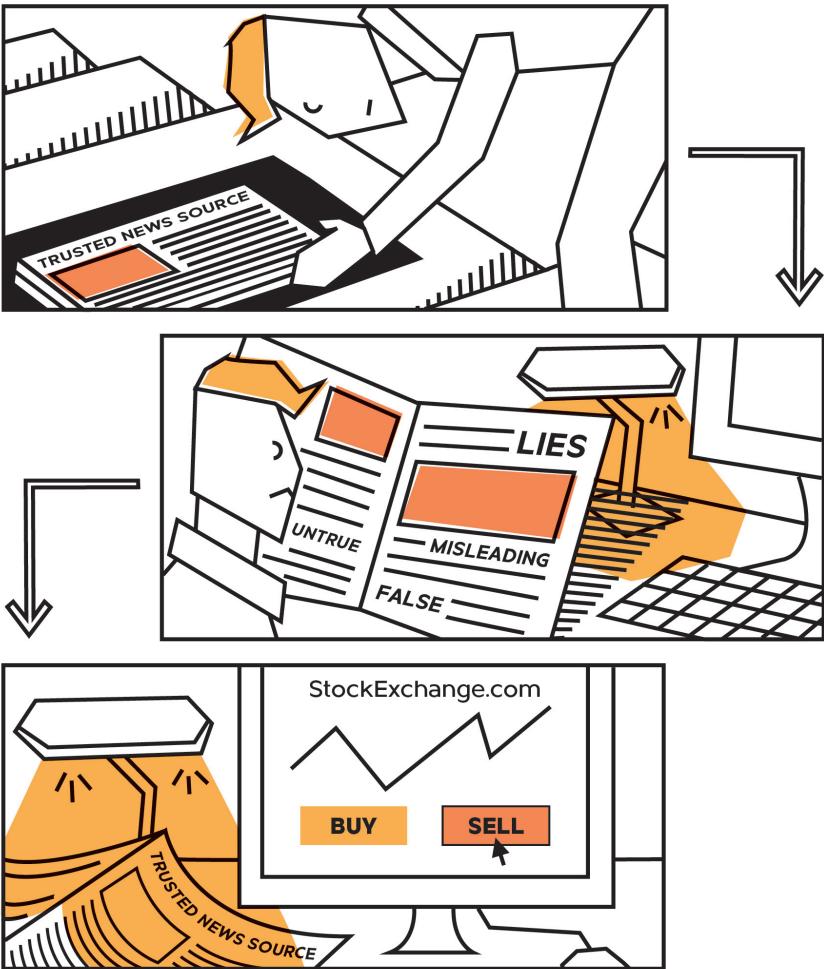
The next call Damien made was to ExchangeWatch, the newswire service that published the release in the first place. If the release didn’t originate internally, ExchangeWatch was the prime suspect. The phone rang several times before Amy Miller, the CEO and Damien’s primary point of contact there, picked up.

“We’re figuring it out, and are five minutes from publishing a public retraction. We don’t know how the release was authorized or by whom—no one seems to be able to recall ever seeing it cross their desks, but we’re going to figure this out—”

“Amy—”

“I’m waiting for the IT team to come back with an audit of all outgoing releases so we can figure out who sent this. I should have that on my desk in the next ten minutes—”

“Amy—”



"Damien, I know, we're doing our best over here, but you have to give me a few more minutes. I know that we should have called you as soon as we realized what happened, but we're trying to figure it out. We run a tight ship over here, Damien, we don't falsify the news, that's for other outlets to do, but that's not the business we're in."

"Amy!"

"Damien, I've got to run and take care of this," she said, "I'll call you back when I have more information." She hung up.

Even if someone over at ExchangeWatch sent out the release, Damien sensed that there was something else going on. How did the press get a comment from

his team if no one could have received a call? Regardless, the first thing he needed to do was to slow down the spinning of the story before the whole thing got out of hand. He sat back down at his computer to type up a public statement, but before he could start, he was interrupted by a knock. His assistant stood in the doorway with a furrowed brow.

“They’ve called an emergency meeting in the executive suite.”

“Right. Thank you.”

Damien pushed away from his desk, grabbed his laptop, and made a beeline for the elevators. Inside, he pressed the upper right button for the executive suite, and the car began to climb through the floors. Before the elevator arrived on the top floor, the phone in Damien’s breast pocket vibrated again. He pulled it out to find a text from Amy.

“We found him.”

CHAPTER 2

PEOPLE, NOT COMPUTERS

REBECCA COZIED INTO HER LIVING ROOM COUCH with a cup of coffee unfurling morning paper across her lap. The cover story, which read “One-two punch: what we know about the hack of CapitalCorp and ExchangeWatch,” provided little more than a timeline, and some speculation about the nature of the attack that had occurred a few days prior.

It was apparent from the article that ExchangeWatch had been very forthright with the press. The article’s author had included some interview excerpts with ExchangeWatch’s higher-ups, including their IT director who had gone so far as to name the poor editor whose account was appropriated for the attack. CapitalCorp, on the other hand, was extremely reticent about their side of the story. In fact, the reporter had only eeked two statements out of the corporate behemoth. The first being a bland declaration from the communications department that the organization was “looking into any impropriety” on their side in cooperation with federal agencies, while the second was a carefully worded proclamation from the company’s CEO about the spurious nature of the claims and his dedication to his firm and family.

While the article was sparse on details, the attack itself was of interest to Rebecca. When people think about cyber-attacks, they often imagine some hooded basement dweller coding their way into a network’s hardware or software systems,

but this attack clearly showed how an attacker could use something like a press release to ‘hack’ people’s perceptions instead. As Bruce Schneier, one of the fathers of modern cryptography once wrote, “Only amateurs attack machines. Professionals target people,”⁴ an adage Rebecca often found herself repeating like a mantra in polite conversation.



However, as Rebecca recalled, this attack was not the first of its kind. Rebecca remembered a story her daughter, Sarah, had told about a similar attack against a company called Emulex Corp. in 2000. A fake press release caused the markets to lose confidence in the firm, sending the company’s stock into freefall. However, once the press release was revealed to be fake, the stock price rebounded, and federal agents caught the perpetrator twenty-four hours later. As Sarah put it, “it was a smart attack, playing the public like that, but the guy was an idiot for not covering his tracks well enough.”

Realizing that she hadn’t spoken to Sarah since before the attack, Rebecca got up from her perch in the living room to give Sarah a call. Sarah worked as a security

analyst for CapitalCorp and had likely spent most of the past few days and nights poring over audit reports to make sure her team didn't miss anything. Rebecca quietly hoped that she'd be able to get a bit more information out of her daughter than what the reporters had managed to dredge up. However, before she could finish dialing her daughter's number, an incoming call came through from an unknown caller. Rebecca picked up and put the receiver to her ear.

"Rebecca Saxon?" A baritone voice asked across the line.

"Yes, this is she. Who is this?"

"Hello Ma'am, this is agent David Kepler of the FBI's Cyber Division. I was given your contact information by one of my colleagues, agent Toni Anette, with whom I believe you're acquainted?"

"Yes, I've done a fair amount of work for Toni in the past. What's going on today, agent?"

"We're currently conducting an investigation, and I was looking for a forensic psychologist with clearance who's done work on cybercrimes in the past. Agent Anette said that you were the best."

"Is this about the ExchangeWatch hack?"

"I'm not at liberty to say, ma'am. If you're willing and able to help us out, I can give you a full briefing once you commit."

"Of course," Rebecca said, familiar with this annoying formality. "When do you need me?"

"Today, if possible."

"Right." She said, scanning through a list of meetings she had scheduled in her day planner, "Let me see if I can move some things around at the lab. It's dissertation research season, so this isn't the greatest time. Where are you calling from?"

"I'm in the New York office. You know where that is?"

"Yes, I'm familiar with it."

"Great. They'll be expecting you downstairs. I'll brief you when you arrive."

Rebecca drank the last sip of her now cold coffee and set to work calling her Ph.D. students to inform them that she wasn't going to make it into the lab. Once finished, she got into her car and drove off to Katonah Station to catch the Metro North into the city. She arrived at the Javits federal building two hours later.

Rebecca walked through the revolving door at the foot of the building into the lobby and checked in at the front desk. Pushing through the turnstiles, she made

her way to the elevator bay and ascended through the building to the FBI field office on the 23rd floor. When the doors opened, a sharply dressed, middle-aged man in a gray suit was waiting. It was agent Kepler.

“Ms. Saxon?” Kepler asked.

“Agent Kepler,” Rebecca said warmly, extending her hand. “Please, call me Rebecca. Only my undergrads use my last name.”

“Pleasure to meet you, Rebecca. Thanks for coming down here on such short notice. Let’s not waste any of your time. If you’ll follow me.”

Agent Kepler swiped his ID on a small keycard reader and led Rebecca through two large glass doors which separated the elevator bay from the rest of the floor. Passing through rows of desks, Kepler brought Rebecca down a hallway to a small Spartan office in the middle of the floor with a desk, a phone, and three chairs. There was a mirror at the back of the room, and a small security camera attached to the ceiling in one of the corners.

“This is one of our interview rooms,” Kepler said as he watched Rebecca take stock of the space. “Sometimes people call them interrogation rooms, but I find ‘interview’ to be a friendlier word. Please take a seat and get comfortable. Can I get you anything to drink before we start? Water or coffee?”

“Water would be great, thank you,” Rebecca said, sitting down in one of the chairs.

Kepler nodded and disappeared down the hall, returning shortly with a glass of water for Rebecca and cup of coffee for himself.

“Here you go,” he said, placing the glass of water in front of Rebecca, and then settling into the seat across from her. Rebecca took a sip of water and, putting down the glass, noticed Kepler scrutinizing her for a split second. Kepler quickly became self-aware and broke his gaze to take a swig from his mug.

“What is it?” Rebecca asked.

He choked down the coffee before he spoke. “I’m sorry. Very few of the experts I’ve brought have resumés as impressive as yours,” he paused for a second, thinking about his words, “...and only a fraction of them have been women.”

Rebecca sighed heavily, “I appreciate that, agent. Despite women’s role in pioneering computing, for the past few decades, it’s largely been a boy’s club, and that includes research around the topic area. It’s an unfortunate fact about this field, as is true with many others, that I needed to stand a couple of heads above my male

colleagues for them to see me at eye level.”

Kepler gave a solemn nod. “Did you always know that this is what you wanted to do?” he asked.

“Hardly,” Rebecca said, “In high-school, I thought I wanted to become an economist. I had always been interested in understanding how people made decisions and believed that following in the footsteps of economists, you know, those who purported to know about such things, would provide me with the insights I was looking for. When I got to college, I took an intro econ class, and I remember walking out of the second or third lecture with a sour taste in my mouth.”

“Bad lecturer?” Kepler joked.

“No, he was quite good, but I couldn’t reconcile what he was teaching with this intuition that I had. Like every other classical economist out there he was steadfast about modeling people’s decisions as if human brains were like simple computers. He said that the “right” way to think about people was as calculating, unemotional maximizers and that didn’t sit well with me. I had seen enough of my friends wake up with hangovers too many times to believe that they were doing a good job of maximizing their utility, and after watching my professor stumble out of the faculty bar one night, I was certain he wasn’t that good at it either.”

Kepler chuckled, “I can see what you mean. But, you’re not an economist, so how did the switch happen?”

“Well, I caught wind of a lecture from a high-school friend of mine at Stanford that I shouldn’t miss. She said it would change my worldview. I was skeptical, but I went anyway. My friend ended up being right, and the professor who gave the lecture turned out to be this guy named Amos Tversky.”

“You mean, like Kahneman and Tversky?”

“So, you’re familiar?”

“I read Kahneman’s book a few years ago. I thought it was great— incredibly insightful.”

“Those insights, the ones that Kahneman and Tversky developed, were what changed everything for me. The next semester I switched my major to mathematical psychology, which more or less combined what I appreciated about the practice of economics with these new insights.”

“Why switch everything up like that instead of just integrating what you were learning into economics?”



A BRIEF HISTORY OF BEHAVIORAL ECONOMICS

“You know and I know that we do not live in a world of Econ. We live in a world of humans.”⁵

—RICHARD THALER

Such were the feelings of a few bold academics (among them, Daniel Kahneman, Amos Tversky et al.) who started noticing flaws in traditional economic models in the 1960s and 70s. Building off of the work of the pioneering Cognitive Scientist Herbert Simon, these economists birthed the field that would later be known as behavioral economics. Behavioral economics, in Thaler’s words, “is not a different discipline; [rather,] it is economics done with strong injections of good psychology and other social sciences.⁶

Today, academics continue to use these early theories to think through ways to help real humans. In practice, taking a behavioral approach means beginning with the propositions that context matters, that awareness does not guarantee action, that we all have predictable biases, and that both monetary and non-monetary costs and benefits drive decisions. It means focusing less on how people should act, how we expect them to act, or how they intend to act, and more on how they actually act.

The behavioral approach generates new ways to examine problems, particularly when focusing on bottlenecks, or specific features of the decision or action context that can affect behavior. With the right tools, identifying those bottlenecks can help to derive fresh and compelling solutions.



“Because I began to fundamentally disagree with some of the major assumptions classical economists made. The rational actor model that economists espoused for years was missing important considerations.

Kahneman and Tversky showed that peoples’ cognition was limited, which caused people to exhibit systematic biases that would, at times, lead them to act in ways that appeared to conflict with their intentions. But, what was most interesting to me was the pernicious effect of context on peoples’ decisions and actions. I realized that depending on the context, I could predict where biases might pop up, and I wanted to understand better how that worked.”

“How did you end up focusing on human-computer interaction?”

“Context, I guess?” She said with a smile, “I had a friend in college who was a brilliant computer scientist. Truly an incredible mind and passionate about it too. One day she came back to my dorm ranting and raving about this Macintosh 128k that the computer science department had purchased for its lab. She said I had to see it, so she grabbed my hand and pulled me across the campus to check it out. It was the first time almost anyone had ever seen a computer with a graphical user interface, but it was immediately apparent to me that everything was about to change. I imagined a universe in which we were all immersed in these computer environments, interacting with ideas and tools in this virtual space, and I asked myself whether all those human biases that Kahneman and Tversky found would end up getting mitigated or amplified in that world. No one else was looking into these questions, so I saw an opportunity to carve out a research niche for myself. I had a few published articles under my belt when I graduated, and by the time I finished my Ph.D. the world wide web was a brand new thing—the whole space opened up. Fast forward a couple of decades, and here I am with my lab, continuing the work I started in the mid-80’s.”

“Well, it looks like you managed to figure out how to keep yourself in high demand indefinitely. It also appears that we never were able to figure out how to use computers to get around people’s cognitive failures.”

“If we’re the ones operating those machines, there will always be human failures. This I can promise you.”

“Which brings us to the task at hand. Your intuitions were right about why I called you in today. We’re in the process of investigating the ExchangeWatch and Capital-Corp hacks and wanted to get your thoughts about the decisions and actions of

some of the key players, with the hope that you can help provide some recommendations about how to avoid these sorts of missteps in the future.”

Kepler briefed Rebecca on the FBI’s current understanding of the situation. The attack vector had clearly been through the compromised account of an Exchange-Watch editor named Peter Frank, but beyond that, it had been very hard for the FBI to unearth any additional information on the motive or the bad actor. Whoever committed the hack had covered their tracks pretty well. Kepler also informed Rebecca that while CapitalCorp had said that they were working in cooperation with the FBI, in reality, they had been quite cagey about providing any data that could potentially support the investigation.

“All we have are bits and pieces, none of which gives a healthy perspective on what happened. So, we have to start at the very beginning. Peter Frank is on his way to the office now, and I’d like for you to interview him to see if you can glean any helpful information about how this all took place, and where we might want to start looking next. You up for it?”

“I’ll help any way that I can,” said Rebecca. “But first, what can you tell me about Peter Frank?”

CHAPTER 3

TAKE ME OUT TO THE BALL GAME

AN AGENT ESCORTED PETER FRANK into the interview room and promptly disappeared to get Peter something to drink. By Rebecca's estimation, Peter couldn't have been more than twenty-three. He was dressed casually in jeans and a flannel and looked intimidated in the presence of the FBI agents.

Kepler stood to shake Peter's hand. "You find the building alright?"

"It was easy enough," Peter said.

Kepler introduced Rebecca as his colleague and offered Peter a seat at the table. "It must have been a crazy few days for you."

The door opened, and the agent who had disappeared returned with a glass of water. He placed the glass in front of Peter and left again, closing the door behind him. Peter picked up the glass with two shaky hands and took a sip, spilling some on the table. Peter scrambled to mop up the water with his sleeve. "Sorry, I think I'm just a bit nervous."

"It's ok," said Rebecca. "This is your first time in an FBI office, right? It's no big deal. Do this with me, just take a deep breath." Rebecca sat up straight in her chair, and lifting her chin with her eyes closed, took a deep breath in and out through her nose. "Give it a try with me."

Peter nodded and then mimicked Rebecca. Sitting tall in his seat, he took a few deep breaths with his eyes closed. He opened his eyes and looked at Rebecca with a timid smile.

“Feeling a little bit better?” Rebecca asked.

“Yes, a bit. Thank you.”

Rebecca gave him a warm smile. “Before we get started, would you like to tell us a little about yourself? Sometimes I find that’s an easier place to start.”

“Sure,” Peter said.

Peter gave them a truncated life story, explaining that he had grown up in a small suburb of Boston, went to college thirty minutes from his parent’s house, and left New England to take a job in the big city. As Peter spoke, Rebecca watched as the anxiety left his shoulders. She looked to Kepler and gave a small nod letting him know he could begin with his questions.

“So, Peter, are you ready to answer a few questions for me?” Kepler said.

“Yes,” Peter said. “I’m ready.”

Kepler nodded, “My forensic team took a look at your computer, and I have to say it was ugly. They found a bunch of malware, most of which was relatively benign, but one, in particular, was not. Do you happen to remember what you were doing with your computer on September twenty-ninth of this past year?”

“What do you mean?” Peter asked.

“The malware we’re concerned with was installed on the twenty-ninth of September, and it’s a pretty mean bug. By exploiting a known vulnerability in your computer’s operating system, the malware was able to break into your stored login information, including passwords, and transmit that information to some currently unknown third party. We believe that is how they were able to obtain your login credentials. Do you remember downloading something that you shouldn’t have? Going to a website you should have avoided? Using an unsecured USB key? Anything like that?”

Peter sat there for a moment thinking. His eyes scanned left and right as he rummaged through his memory. All of a sudden, he stopped and looked up. “I think I might know what happened. Let me check something real quick.” Peter reached into his pocket and pulled out his cell phone and began to scan through his calendar. “Yeah, I think I remember that day.” Peter put his phone back into his pocket. “I stayed late at work because I had a big assignment due in the morning. I remember being frustrated about it because there was a Red Sox game that I wanted to watch, but I couldn’t leave the office and watch at home, so I decided that I’d try to stream the game on my computer. It was going to be like twenty bucks to stream from the



THE AFFECT HEURISTIC AND BAD LINKS

Research has shown that people sometimes judge the risk of taking a particular action, not based on any calculated risk, but instead on how they feel about the decision or action. If they feel positive about the outcome of that action, they are more likely to judge the risks of following through as low and the benefits high, while the opposite is true if they feel negative about it.⁹

Behavioral scientists have found this appeal to affect (emotion) often determines the perceptions of risks in significant ways. Calculating risk can sometimes be complicated, so to simplify things, people often use a shortcut called the **AFFECT HEURISTIC**. In other words, instead of weighing all the potential costs and benefits of a particular action or decision, people rely on their emotion, or "go with their gut," when considering whether to follow through on a potentially risky activity.

To help illustrate this, consider an intuitive (and perhaps unsettling) example of the affect heuristic in health. Cigarette advertising is sometimes designed to increase the positive emotions associated with smoking by presenting images of attractive people happily puffing away. By deliberately associating smoking with positive imagery, advertisers may have the effect of significantly diminishing perceptions of smoking's substantial risks.¹⁰

In the case of Peter, who loved watching the Red Sox play, the positive affect associated with his hometown team caused him to misjudge the benefits and risks related to streaming the game and downloading malware infested software. Being able to watch the game can only be a good thing, right?

While it may be difficult to override Peter's die-hard Red Sox fandom, solutions in this space should consider concretizing his risks and making his potential costs more salient by putting vivid information about specific malware consequences into browser warnings or search results. Instead of letting Peter leave the decision up to his gut, good user interface design could provide consequential information to Peter when he needs it most.



MLB website, so I looked for a free site instead.”

“Did you manage to find a site?” said Kepler.

“Yeah. I think I looked through five or six search pages before I found one that worked.”

Rebecca wrote a note down on a pad of paper in front of her. “When you say ‘it worked,’ do you mean it worked immediately? There were no other steps besides going to the page and starting the stream?”

“It wasn’t that simple. I remember that it didn’t work at first. When I clicked to start streaming the game I got a pop-up saying I had to update the media player.”^{ii,7}

“And?” Rebecca asked.

Peter looked down at the desk, “I mean, yeah, I clicked on it, but I wanted to watch the game. I didn’t think something like this—” Peter took a deep breath. “I didn’t think something like this would happen.”

Peter looked despondent. Rebecca reached across the table and put her hand on Peter’s arm. “It’s ok, Peter. This kind of stuff would happen to my daughter all the time. I remember some years ago having to clear my computer of some gross adware because she had her friends over and they were using my computer to try and find an illegal stream of some movie that had just been released in the theater. Nine of the ten sites they tried didn’t yield anything, but on the tenth one, found the movie and pick up a whole host of infections in the process.”

“Your daughter sounds like the conscientious type,” Kepler joked.

“She’s a security analyst now, so she knows better. But, Peter, my point is that you’re not the criminal here, and you’re doing what you can to fix the situation by talking with us. You alright?”

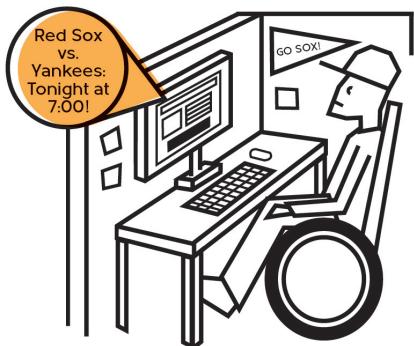
“Yeah, I’m ok,” Peter said. “It’s just frustrating.”

“I understand,” Rebecca said. “Let’s change the subject a bit. So, you like the Red Sox?”

Kepler turned to Rebecca. “Why is this relevant?”

“It’s relevant. Just let him answer the question.”

ⁱⁱ In one study, users were 28 times more likely to be infected by malware when visiting piracy sites than compared to a control group comprised of legal sites. Another report found that as many as 50% of the video overlay ads on free livestreaming websites are malicious. Risk IQ Digital Bait report: <https://media.gractions.com/314A5A5A9ABBBC5E3BD824CF47C46EF4B9D3A76/0f03d298-aedf-49a5-84dc-9bf6a27d91ff.pdf>; KU Leuven-iMinds & Stony Brook Study: <https://www.kuleuven.be/english/news/2016/malware-data-theft-and-scams-researchers-expose-risks-of-free-livestreaming-websites>



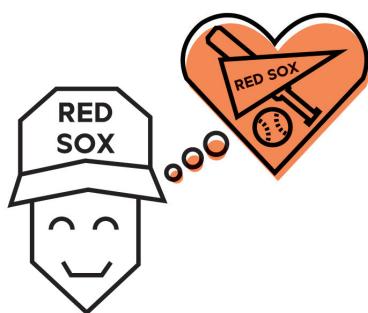
6:55 PM



6:56 PM



6:57 PM

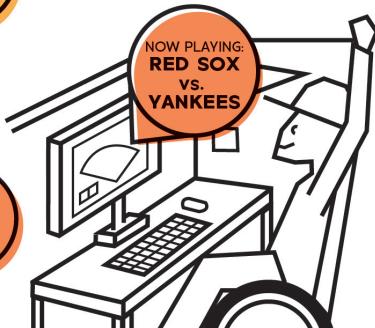


6:58 PM



6:59 PM

7:00 PM



///

WARNINGS, HABITUATION, AND GENERALIZABILITY

A cross the board, getting users to be attentive and adhere to security warnings can be quite difficult.^{11,12,13} This is concerning because warnings represent one of the last lines of defense in protecting computer users from malicious attacks and vulnerabilities. However, users, like Peter, are confronted with various sorts of warnings on a regular basis, many of which have few immediate consequence if ignored, and some simply being false positives.¹⁴ This is a problem. Psychologists have been studying how people react to repeated stimuli for many years and have found that over time people's psychological and emotional response to those stimuli decreases, a process called habituation.¹⁵ In the context of warnings, habituation can help explain why people tend to ignore and click through warnings automatically.¹⁶

HABITUATION can be even more problematic because of the similarities of various user interaction paradigms across different types of warnings. For instance, browser-based SSL warnings and malware warnings look relatively similar and require similar actions from the user to either adhere or ignore them. Habituation to one warning (SSL) can generalize across to other similar looking warnings (malware, etc.), and even warnings and other prompts that share similar user interaction elements (e.g. update prompts).¹⁷ However, the more pernicious problem is not merely that users are generalizing their habituated actions (clicking through) across different warning types, but that they also may be generalizing the perceived risk associated with the warnings they ignore across different warning types, despite the fact that some threats are much more significant than others.

One way that we might fix this problem is to build warnings that don't incorporate familiar UI elements and require the user to do different sorts of actions if they wanted to click through. For instance, some researchers have examined how polymorphic warnings or those that change shape and color can improve adherence to warnings.

///

"Uh, yeah." Peter said. "I love the Red Sox. I grew up in suburban Massachusetts, so it's basically in my blood."

"Baseball generally?"

"I used to play in high school. I like baseball a lot."

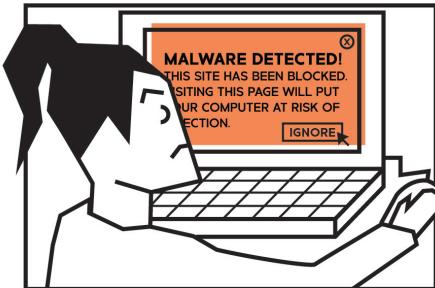
"You watch many games?"

"Probably close to a hundred fifty games a season. I watch almost every Red Sox game, and the one on the twenty-ninth was a big one. The Red Sox were playing the Yankees. I never miss those. That's part of the reason I wanted to find a way to stream it."

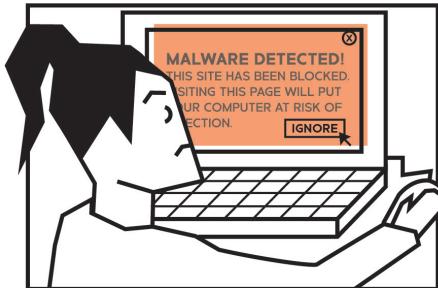
"Did you think about the risks involved in streaming the game illegally?"

"I mean, I guess, but I wasn't worried about it. I knew that I might get something on my computer, but I just wanted to watch the game."

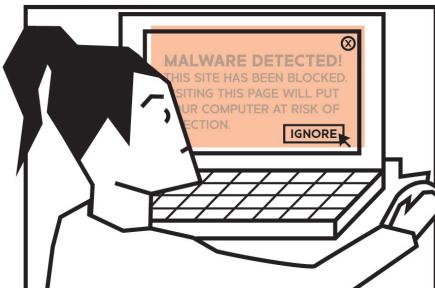
First Time



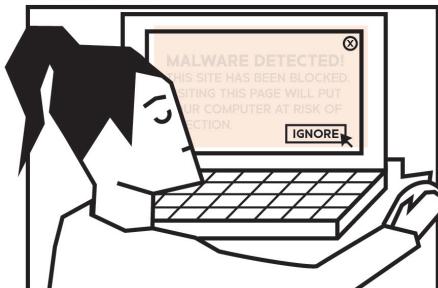
Second Time



Third Time



Fourth Time



Rebecca wrote a note on a notepad in front of her. "Before the website loaded, did you get any warnings? Anything letting you know that there might be malware on the site?"

Peter thought about it for a second. "Now that you mention it, yeah, I did."

"Do you remember what the warning said?"

"No, not really, but I don't trust those warnings. Warnings pop up all the time for



THE CHOICE ARCHITECTURE OF UPDATES

every decision—from choosing a type of toothpaste at the supermarket to deciding which field of cancer research to fund—entails contextual features that influence the resulting behavior.

How many types of toothpaste are there? Where on the shelf is the toothpaste, eye level or at the bottom? Which tube of toothpaste is next to the one you were initially considering? How much time do you have to make the decision? These questions represent a tiny slice of what a behavioral scientist might ask when examining a decision's **CHOICE ARCHITECTURE**. Choice architecture is a term coined by Cass Sunstein and Richard Thaler to describe how the design of a particular choice context (e.g. grocery stores, update interfaces, etc.) can influence people's decisions and actions in predictable ways.¹⁸

The decision of whether or not to update a computer system is no exception to Sunstein and Thaler's framework—in this case, the presentation of Peter's choice certainly contributed to his failure to install a much-needed security patch. More specifically, updates often require a quick and on-the-spot decision: Do I install now or later? If people choose to defer, the update system will ask if they'd prefer a reminder “tomorrow,” “tonight” or “later,” which at first glance may allude to a particular time, but that time may not be precise enough for the user to follow through. Moreover, update systems often present this decision when the user is short on time or attention, further increasing the incentive to defer. Such a situation isn't doing Peter any favors, especially if the update prompt comes when he's least likely to stop whatever he's doing.

A wise “choice architect” might consider helping the user make a concrete plan to update by providing more information about the particulars of the patch such as its purpose and expected install time. Additionally, it might be prudent to ask users to schedule a specific date and time so they can plan and commit to updating in the future. One could also just remove the choice altogether and push updates automatically to the user.



legitimate sites. I was on a State Department site a few months back before I took a trip and I got a browser warning. It's just hard to know what's real.”²

“Was the warning a malware warning or an SSL warning?”

“What’s the difference?”

“They’re different, but if you don’t know the difference, we can move on. How frequently do you update software on your computer?”

“Every once in a while, but why does that matter?”

“Well –”

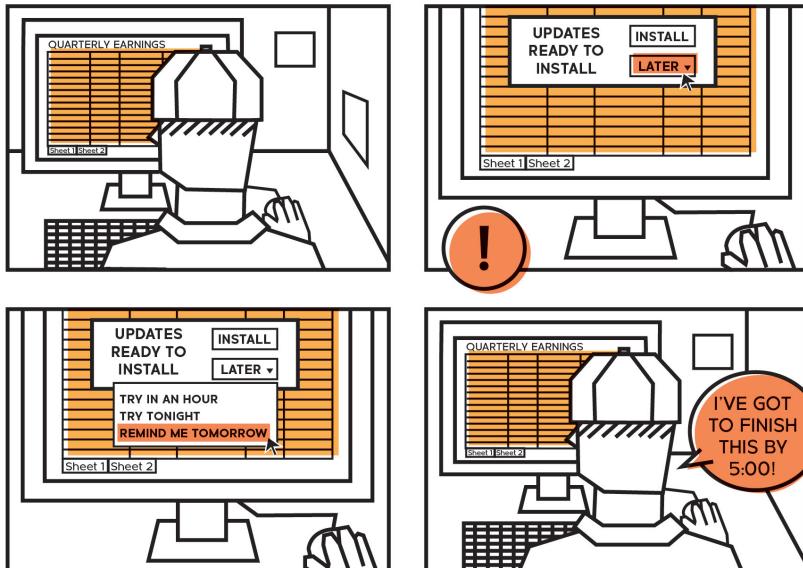
“The malware that was used to compromise your login information exploited a flaw in the operating system software,” Kepler said. “However, the people that built the software found out about the error they embedded almost six months ago, and released a patch a few months after that. You could have avoided all this had you been regularly updating your system software.”

“Can you blame me? My computer asks me to install updates at the worst times. Whenever I’m prompted to install an update, I’m usually in the middle of something, so when it asks me if I want to update now or later, I always choose later. That doesn’t happen to you?”

“When did you last install an update?” Rebecca asked.

“God, I don’t even know. It must have been months ago, I guess.”

“Do you remember how you decided it was the right time to install the update?”





THINKING FAST AND THE INTERNET

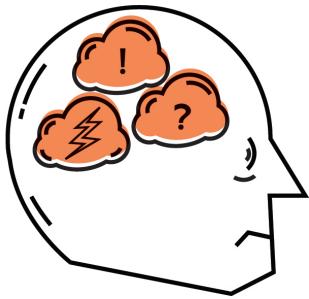
In his award-winning book “Thinking, Fast and Slow,” Daniel Kahneman builds on an idea from psychologists Stanovich and West, that our decisions result from two systems at work in our brains: “**SYSTEM 1**” and “**SYSTEM 2**.¹⁹ System 1 is fast-thinking, operates automatically and intuitively with little or no effort and no sense of voluntary control. System 2 is slower and requires conscious effort. System 1 dominates many of the numerous decisions we make every day. We use System 2 less frequently because it demands more effort.

System 1 is at work when you can sense anger from someone’s tone, or when you decide to put on shoes every morning before you go outside. System 2 is at work when you successfully fill out and file a tax form each year. Our reliance on System 1 surprises many despite its necessity. If we had to think about every tiny decision consciously, we’d be paralyzed by choice and never leave the house.

However, our reliance on System 1 can negatively affect our decision-making. Instead of considering many decisions with the deliberateness that they deserve, we instead use mental shortcuts that generally serve us well, but can sometimes cause us to misjudge risks and likelihoods,²⁰ be inattentive to small details,²¹ plan poorly,²² or make us over-confident in our abilities.

Our computers, smartphones and the like, have helped shape a world dominated by speed. Consequently, we are foisted into a world of System 1 thinking where slowing down is seen as an inconvenience.²³ This wouldn’t be a problem if computer algorithms could perfectly predict when a website poses a threat, or when an email is actually a phishing attack, but so long as we can’t outsource our risk assessment to computers, people will continue to err if they continue to operate in automatic mode.





Automatic Thinking



Fast



Unconscious



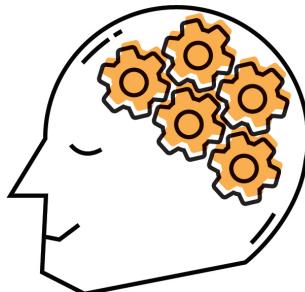
Automatic



Everyday Decisions



Error Prone



Deliberative Thinking



Slow



Conscious



Effortful



Complex Decisions



Reliable

"I guess, sometimes I just want to take a break from work? I can let the computer sit for thirty minutes and walk away. That's probably what happened."

Rebecca had finished her questions, but Kepler continued with Peter for another fifteen minutes, asking him about his relationship with CapitalCorp and their personnel. CapitalCorp was not one of the organizations that Peter regularly worked with as a junior editor, but he said he had pushed out a couple of press releases for them over the past year when one of his colleagues was out of the office. At the end of the interview, Kepler called in another agent to escort Peter to the elevator bay.

Kepler leaned back into his seat, hands cupping the back of his head. "All of it felt mundane. The guy doesn't update his computer, blasts through a browser warning,

clicks on some bad links and gets his computer infected. It could have happened to anyone.”

“Yeah, but it didn’t,” Rebecca said. “It happened to him.”

“So what do you think?”

Rebecca sat with her elbow on the desk; fist pressed into her cheek. “It’s just strange. His emotions about the game got the best of him, and he made a riskier decision than he might have otherwise, but I just don’t get how that could have kicked the whole thing off. It’s not as if the malware targeted him in particular; anyone could have stumbled onto that page. Whoever set the trap just got lucky.”

“Maybe.” Kepler looked up at the ceiling as he thought. “It’s crazy to consider how all of this could have been avoided had he just kept his software up to date.”

“That always gets me. Better design could help fix a lot of these behaviors, but the people who are concerned with creating the user experience always think that you’re going to piss off the user if you’re too focused on security. You remember what I was saying before about whether computers could help mitigate people’s biases?”

“You mean about the Mac 128k?”

“The people who design computer hardware and software would rather help people make faster decisions than slower ones. It’s always about more processing power, faster internet speeds, and more responsive interface designs, but that comes at a cost. There are some circumstances when people should be acting deliberatively and thinking critically about a risk they’re about to take, but designers have made it far too easy to blast through a warning and dismiss an update prompt. When people think about this stuff too quickly, that’s when they take mental shortcuts and get themselves into a compromised situation.”²⁴

“You’re talking about Kahneman again? System 1, System 2?”

“Basically. I think solving these sorts of problems is about figuring out when you need to slow down the user, get them to consider an action before they take it. Otherwise, we’ll just have more incidents like this pop up,” Rebecca said. “Anyway, agent, are we finished for today? I need to get going. I have some work to get done for my students.”

“Yeah, we’re done. Thanks for coming down here today,” Kepler said. “I’d like for you to join a few other interviews once I get them set up.”

“Just let me know,” Rebecca said.

Kepler walked Rebecca back to the elevator bay. As Rebecca descended into the lobby and out the front door of the building, she still felt unsettled about Peter Frank. It was as though there was still some critical missing piece just beyond her grasp. Why had Peter Frank been targeted?

CHAPTER 4

D33PTH0UGH1

[7/14/2016 16:25] <D33pTh0ugh1> update?

[7/14/2016 16:25] <NerfHerder> I found it for you

[7/14/2016 16:25] <D33pTh0ugh1> how many?

[7/14/2016 16:25] <NerfHerder> four

[7/14/2016 16:25] <NerfHerder> there were others but no logins

[7/14/2016 16:25] <D33pTh0ugh1> perf

[7/14/2016 16:26] <NerfHerder> u owe me

[7/14/2016 16:26] <D33pTh0ugh1> I'll put the .4 btc into your wallet

[7/14/2016 16:26] <NerfHerder> .5

[7/14/2016 16:26] <D33pTh0ugh1> we agreed on .1 per

[7/14/2016 16:26] <NerfHerder> change of plans

[7/14/2016 16:26] <D33pTh0ugh1> we had an agreement

[7/14/2016 16:27] <NerfHerder> do you want these

[7/14/2016 16:27] <D33pTh0ugh1> .4

[7/14/2016 16:27] <NerfHerder> do you want these

[7/14/2016 16:27] <D33pTh0ugh1> fine

[7/14/2016 16:27] <NerfHerder> where do you want me to drop the logins?

[7/14/2016 16:27] <D33pTh0ugh1> ssh to 104.139.245.40 whats your incoming

[7/14/2016 16:27] <NerfHerder> 189.141.39.57

[7/14/2016 16:27] <D33pTh0ugh1> pw 01100001 01110011 01110011

[7/14/2016 16:28] <NerfHerder> really? ill send along when i get the deposit

[7/14/2016 16:29] <D33pTh0ugh1> sending now



THE FILE FROM NERFHERDER APPEARED in the server less than a minute after the payment went through. D33pTh0ugh1 opened the file and, scanning through its contents, quickly discovered that the information she was ultimately looking for wasn't there. Instead of finding the usernames and passwords for the four employees' ExchangeWatch accounts, NerfHerder delivered a constellation of various logins for other web services, social media accounts, and web merchants.

And that was the cardinal truth: while deep web smugglers like NerfHerder could find most anything after looking in the right places, finding *most anything* was more common than finding *anything*, so NerfHerder's inability to turn up all of the specific login credentials wasn't unexpected.

Even without the ExchangeWatch login info, the data dump still proved helpful. Hidden among the social media passwords and merchant logins were enough clues for D33pTh0ugh1 to piece together what she needed.

The first step was to figure out the targets' ExchangeWatch usernames. However, this took relatively little effort. The vast majority of enterprise services don't require employees to generate usernames for themselves. Instead, often an organization's administrator sets up employee accounts using the employee's corporate email as their username. Using emails as usernames might be easier for the employees to remember, but it also takes a lot of the guesswork out of the hack. D33pTh0ugh1 did a web search for “@exchangewatch.com email” and combed through the results to see if she could figure out how to reconstruct the account logins for the four employees.

The search results turned up a few examples of ExchangeWatch emails. The first was on an email list maintained by some fraternity alumni group on a publicly accessible web page. The second one she found was at the bottom of a blog post titled, “How to write press releases to get traction,” authored by an ExchangeWatch employee, while the third was on a year-old list of emails compiled for attendees of a public relations conference. D33pTh0ugh1 found that each of the ExchangeWatch emails followed the same format of *first initial last name @ exchangewatch.com* and was quickly able to reconstruct the usernames of the four targets.



INCOMPLETE RULES OF THUMB

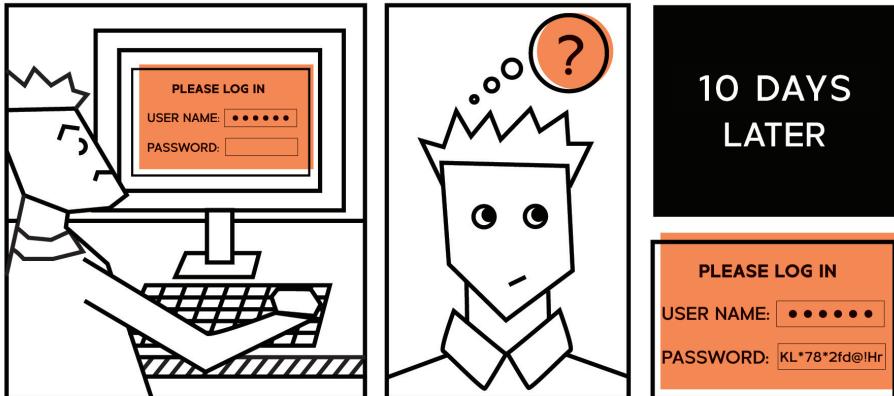
The now all-too-familiar secure password rules (e.g. including at least one upper case and one lower case character, one number, one special character, etc.) were originally devised to help guide users to create passwords that were computationally harder to crack using brute force techniques. However, these rules were intended to be applied to randomly generated passwords, which is problematic because human beings struggle at both creating and remembering random strings of characters (e.g. letters, numbers, symbols, etc.). Instead, people end up applying these rules in ways that are *systematic* and *predictable*, undermining their security in the process.

For instance, one could imagine someone adhering to the general rules of thumb for strong passwords while still making a password that's easy to guess. The user might construct a password first by considering a password or phrase that is easy for them to remember, such as "randomness," and then applying the strong password rules afterward. For instance, they may change the "r" to R, the "o" to a zero, and the "s" at the end to a "\$" so that they end up with "Rand0mnes\$" instead. Entering in a password like "Rand0mnes\$," users might get lulled into a false sense of security, believing that they had followed the rules of thumb well, reinforced by the fact that the user interface provided feedback that the password was "strong." However, a password like "Rand0mnes\$" would be easy to crack because of the formulaic application of the secure password rules. Understanding that people aren't good at doing anything randomly, we must ask: How do we help users generate secure passwords that are also easy to remember?

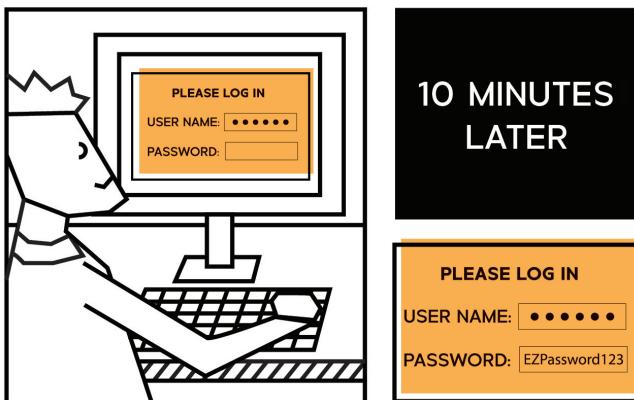
To combat these issues, sites that require authentication should consider enforcing new rules of thumb that make randomness a requirement while still producing easy to remember passwords.²⁵ One stronger heuristic would be to encourage the use of passphrases that leverage randomly selected dictionary words (e.g. "falsefrogpaperbell" or "obesedragonmagentatissue") to nudge users to use longer, more memorable passwords that are simultaneously harder to crack. Additionally, to get around weak password construction, websites may need to increasingly focus on ways to nudge users to adopt two-factor or multifactor authentication.



STRONG PASSWORD: HARD TO REMEMBER



WEAK PASSWORD: EASY TO REMEMBER



The next step was to crack their passwords. While a little trickier than divining usernames, cracking passwords was far from an impossible feat. Besides, D33pTh0ugh1 only needed one to succeed. She pulled up the lists of passwords used by the four targets on other sites and began combing through them. Two of the four targets appeared to be abnormally conscientious about their password choices. Each of their multiple passwords looked like a random assortment of characters, numbers, and symbols. D33pTh0ugh1 knew she could try to use brute force methods to figure out the passwords, but considering the ease with which a diligent IT security manager could use their host intrusion detection or log correlation system to identify this kind of attack, she decided on a more clandestine strategy.

The two other users had not been nearly as careful as their colleagues. It wasn't as if they had just chosen 'password,' but their choices weren't much better than that. One looked like they had done their best to adhere to some of the relatively standard "strong password rules"ⁱⁱⁱ such as using at least eight characters or including numbers and symbols. However, these characteristics were tacked on to the beginnings or ends of easily searchable dictionary words, or used in place of easily recognizable letters. For instance, the password "doggie" had become "Doggie1!" and "pastasalad" had become "P@st@s@l@d*." To the users' credit, despite the formulaic application of the strong password rules, the dictionary words used across the passwords were entirely different. Again, not an insurmountable barrier—but, there was a simpler alternative.

The last employee, one Peter Frank, was by far the least secure of them all. For starters, Peter's social media password had been captured in the keychain and was clearly an artifact from a bygone era. The password read "fenway," was six characters long, and used none of the strong password rules. It was the sort of password that the social media company no longer allows new users to create; yet, here it was. Peter had probably generated that password for himself in middle school and never changed it. The remaining passwords in the keychain were "Fenway11," "Fenway!!," "RedSox1!," and "Redsox!1." Jackpot.

D33pTh0ugh1 wrote down the four password combinations on a piece of paper, and added to them some similar variations, replacing '1's with exclamation points, and selectively capitalizing and lower-casing several letters. Testing the combinations was a simple exercise, but took a bit of time. To make it harder to detect her intrusion, she spaced out the testing of each password over a couple of days, never trying more than one password at a time and always waiting a few hours before a subsequent attempt. She always made sure to use public Wi-Fi and protect herself

ⁱⁱⁱ While strong password rules are intended to get users to build passwords that are harder to crack, they are best applied randomly. However, people have a tendency to use these rules in very predictable ways. Take, for instance, the passwords Behave42 Scienc39 and Noplac33 - long and seemingly complicated passwords that would likely pass most password strength tests. However, a good hacker can easily crack these sorts of passwords. Each of the three passwords above represents a standard password "topology." This particular topology follows the pattern of one uppercase letter (u) five lowercase letters (l) and two digits (d) or "ullllldd" for short, which in one study was found to occur in over 12 percent of cracked passwords. Hackers can use this insight to build software that can crack passwords in smarter, more efficient ways. - Hank Leininger (June 2014) PathWell: Password Topology Histogram Wear-Leveling. Kore Logic. Presentation for BSides Asheville 2014. Retrieved from https://www.korelogic.com/Resources/Presentations/bsidesavl_pathwell_2014-06.pdf on October 20, 2016



STATUS QUO BIAS AND PASSWORDS

Passwords are sticky. Once a user creates and remembers a password, chances are they'll keep using it, which can significantly reduce its security.²⁶ When signing up for a new website or service, users reuse passwords they're already using on other sites. Additionally, if a password must be changed, the user may change one character from a previous password (e.g. password1 to password2). By sticking with one or two different passwords across multiple services, users put themselves at greater risk. If one of those services is compromised, a bad actor may gain access to others, including those that are sensitive (e.g. bank accounts, medical records, etc.)

The psychological phenomenon called status quo bias can help to explain why users stick with and reuse passwords again and again. **STATUS QUO BIAS** is an emotional overweighting of the current status of affairs which decreases the likelihood that people will make changes in a particular context. There are many contexts in which status quo bias can arise, but there are two of particular interest for passwords.

First, when prompted to create a new password, users are confronted with a nearly infinite number of choices, and can become overwhelmed and stick with a password they already use. In fact, status quo bias can intensify when there are a large number of options from which to choose.²⁷

Second, a user's awareness of her own limited memory may nudge her to keep a current password. In creating a new password, users may become concerned that they'll forget a new password that dissimilar to previous passwords.²⁸ Then, by focusing on the costs of switching passwords, (e.g. getting locked out of accounts if they can't remember the new password) users may experience loss aversion. **LOSS AVERSION** is the idea that people weigh losses more than equivalent gains, and are therefore more motivated by costs than they are by benefits.²⁹

We might design around status quo bias by getting users to adopt additional authentication methods (e.g. multifactor authentication) to add extra layers of security to their passwords. Another idea is to build additional function into password interfaces, by for instance, preventing users from choosing passwords that are too similar to older passwords, or by generating passwords on behalf of the user that are easy to remember.



with her VPN, making it much more difficult for IT personnel to figure out who she was and where. She cracked the system on the second day. Sitting in a small cafe in the middle of Manhattan, D33pTh0ugh1 typed “FenWay1!1” into the password bar, and instead of returning with an incorrect password message, the browser continued to load, quickly bringing her to the ExchangeWatch employee landing page. But before she could do anything, a popup materialized on the screen, reminding Peter that he was overdue to set up two-factor authentication for his account. She breathed a sigh of relief, considering the challenge she would have had to face if Peter had been more diligent about his personal security. Guessing a weak password was one thing, but getting past two-factor authentication was a whole other ball of wax, and something she would rather avoid.



THE ELEVATOR DOORS OPENED, and Damien stepped out into the glass and steel atrium that was the executive suite. Two-story windows on either side of the building provided expansive views of both the Hudson and the East Rivers, while a skylight above opened an unobstructed view of pillow clouds against deep blue. Damien, who was preoccupied trying to get Amy on the phone, paid no attention to the beauty of the architecture around him. After receiving her text message, he attempted to reach her twice. Both efforts went straight to voicemail. However, on the third attempt there was a ring, and then Amy’s exasperated voice on the other line.

“You got my text, right?” she asked.

“I did. What’s happening?”

“He was one of ours,” she said. “He denies knowing anything about it, but it came from his account. No idea how that happened. What a goddamned mess.”

“So who was it?” Damien asked.

“One of our young editors, a guy named Peter Frank. To be honest, I think he lacks the stomach to do something like this, but I guess you never know. We’re going to have to put him on administrative leave until this gets sorted.”

“What about the retraction statement?”

“Give me another twenty minutes. There are a few desks it’s going to have to cross before we put it out there.”



THE HASSLE WITH MULTIFACTOR AUTHENTICATION

nnoyances (real or perceived) associated with small tasks—particularly those that involve complex processes or unanticipated steps—often elicit what we call a **HASSLE FACTOR** or minor inconvenience that generates relatively dramatic impacts.

One well-known example of hassle factors that many college applicants face is the Free Application for Federal Student Aid (FAFSA). The eight-page, 100-plus question FAFSA form is an entirely daunting task, and can take a significant amount of time and effort to complete. However, the benefits are well worth it. Not only does completing the FAFSA make students eligible for thousands of dollars in grants and subsidized loans for college, but it also increases the likelihood that they'll attend college the next year. Many families still fail to complete the form.³⁰ To classical economists, this behavior does not make any sense as the benefits that would accrue to families from completing the FAFSA far outweigh the costs associated with filling out the form. But, behavioral scientists recognize that these small hassles can have an outsized effect on human behavior.

In the context of user security, the existence of hassle factors can help to explain the low rates of adoption of two-factor (2FA) and multifactor authentication (MFA) despite the significant security benefits of their use. Specifically, users are often required to opt-in to 2FA and MFA on their own if they want to use it, which means navigating to a website or service's infrequently trafficked security settings page, turning on the service, and then setting it up with a phone or another device on hand. Additionally, users may be wary of the potential hassles associated with needing to use their phone or some other device to authenticate themselves (e.g. “what if I don’t have cell service, my phone dies, or I need to authenticate from a foreign country?”) which also reinforces their perception of hassles involved.

One solution that might help users get around the hassles would be to change the adoption choice from an opt-in to an opt-out by defaulting users into setting up 2FA or MFA when they initially sign up for a web service. Doing so could create the opposite scenario—the small hassles associated with opting-out would likely lead more people to keep the service. Another option might be to provide a single click option or deep links for setting up 2FA or MFA when defaulting users into the service isn't feasible.



“Including mine, thanks.”

“Right.” There was an audible sigh. “I’ll send it over as soon as I can. I have to get back to this.”

Damien hung up and returned the phone to his blazer’s breast pocket. The morning sun shone sideways through the building casting long shadows of the low-slung leather seats and coffee tables in the foyer. He began making the trek to the back of the suite, where a curved white stairwell led to the boardroom. At the top of the stairs, Damien could see Dana, CapitalCorp’s CISO, and James Robinson, the CEO, sitting at the long, oval boardroom table. Another woman was sitting with them whom Damien did not recognize. He wondered if it might be a lawyer. As Damien opened the doors, James looked up and smiled at him.

“Come on in, Damien. We’re just getting started,” James said.



AVAILABILITY AND THE RISKS WE SEE

Amy had assumed the hack against ExchangeWatch had been carried out by a sophisticated hacker, but instead, it was actually caused by a naive employee. While companies invest heavily in external threat preparedness (e.g., virus, worms, Trojans, and bots), when it comes to internal threats (either malicious or unwitting) by trusted agents, many organizations may not be prepared. In fact, 43% of data breaches have been found to come from inside organizations.³¹

One explanation as to why organizations focus so much on external threats when insiders (often unwittingly) cause such a large proportion of breaches has to do with a human bias called the **AVAILABILITY HEURISTIC**. The availability heuristic is a mental shortcut that people use when considering the probability of something happening. Instead of calculating the actual likelihood of an event occurring, people instead estimate probability based on how readily available a given memory of that event happening is. However, what's easy to remember may not be very likely, so this method of estimating likelihood can produce very biased results. News outlets are often implicated in perpetuating this bias because the news primarily functions by making salient relatively rare events—everyday things are by definition not newsworthy. In the context of cybersecurity, the vast majority of news stories that cover cyber-attacks often focus on external threats and malicious hackers, not the mundane mistakes of employees. Because of this, people are less likely to consider insiders as a primary threat vector, instead focusing on the external risks which may not be the greater concern.

However, the bias of availability may not be all bad. By reporting on hacks, news outlets may be helping to change the perspective of individuals and organizations who had potentially underestimated the need for cybersecurity infrastructure. Instead, they may now see cyber threats as a significant, albeit in some cases overstated, risk. While these individuals and organizations may be overestimating the risks they face, being over prepared is far better than being underprepared in the context of cybersecurity.



CHAPTER 5

OPEN ACCESS

KEPLER CALLED REBECCA WHEN she was on her way back to Westchester. “I have an interview on the books for tomorrow with some folks from Exchange-Watch,” he said. “CapitalCorp’s still tight-lipped about the whole thing, and I’m losing confidence that they’ll open up to us. Anyway, you think you can come by?”

By Rebecca’s estimation, this kind of maneuvering from a Fortune 50 like Capital-Corp was not entirely unexpected. She had observed many well-known, well-funded organizations that, after falling victim to a cyberattack, were reluctant to have the Feds (or anyone for that matter) carefully scrutinizing their information security infrastructure, particularly their audit logs. Even the insurance providers she had interviewed often complained that their own clients barred the insurer from accessing information needed to process claims and build useful actuarial models.^{iv}

^{iv} In theory, the growth of the cyber insurance market should offer the security community a robust tool for identifying ‘what works’ in cyber security. After all, if insurers are collecting risk assessment data on each customer before issuing a policy, and collecting incident data each time a company files a claim, then over time there should be robust data to determine the marginal benefits (in risk reduction) of each type of security control (e.g. mandatory 2-factor). In practice, insurers collect shoddy information when selling policies, and often pay claims without gathering any forensic data on the nature of the compromise. Enterprises purchasing insurance will complete required self-assessments based what their internal security practices should be, not necessarily what they are in practice. When processing a claim, insurers often don’t seek—or seek, but don’t require—insured parties to provide robust data about how a hack or breach occurred. Even sophisticated insurers who seek more data from their customer for the purpose of market research to develop their insurance products find most enterprises unwilling to cooperate.

It was understandable. A visible and newsworthy hack can erode public confidence in an organization, which can directly affect the bottom line. It was the leadership's responsibility to protect their shareholders' interests (and their own skin), making it much easier to justify sharing minimal information, if any, about the nature of a hack, particularly when information revealed during an investigation can eventually become public record. However, in Rebecca's experience, while stock prices did often decline in the wake of a hack, they often rebounded within a couple of months.³² The public generally has a short memory for these sorts of things.

As Rebecca's daughter, Sarah, had once said, "A lot of people expect that their personal data isn't secure and that the organizations they 'trust' are likely going to get hacked. People think it's inevitable in the same way that plane crashes are inevitable, but people still fly."^v

The difference, as Rebecca had pointed out, was that airlines are required to document and report back to the FAA about what went wrong so that airlines can fix the planes and prevent that same problem from occurring again.^{vi}

"Maybe that's what's next for the cybersecurity industry," Sarah had said.

"I hope so," thought Rebecca.

She was still thinking about that conversation the next day as she rode the elevator to the FBI's 23rd story field office.

Kepler was waiting for her in the elevator bay again when she arrived. The two of them made the trek back through the field office to the small interview room where they had met with Peter Frank the day before.

"They're not coming in," Kepler said. "The meeting's not called off, but we have to call them."

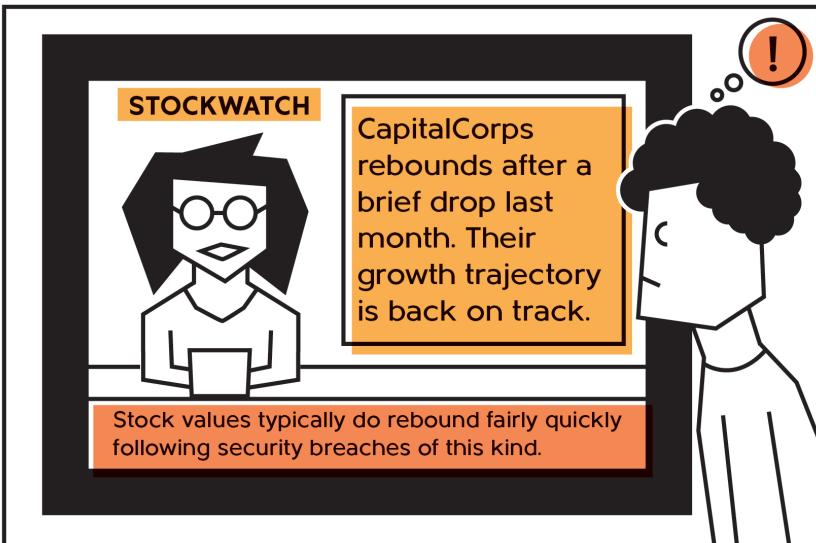
^v According to a Pew Research survey, "Americans have very little confidence that their data will remain private or secure." More specifically, American's lack faith in the organizations charged with managing and securing their data. The survey found that only 6% of adults are "very confident" that government agencies can keep their data private and safe, 6% of adults say the same thing about their activity records concerning landline telephone companies, and only 1% make that claim for social media sites. In spite of these beliefs, "while some Americans have taken modest steps to stem the tide of data collection, few have adopted advanced privacy-enhancing measures."

^{vi} A handful of industry experts has called for the U.S. government to create a new federal agency dedicated to technology policy. Just as new agencies were set up to regulate rapidly proliferating technologies (e.g. cars, planes, mines, telephones, television, etc.), the rise of hyper-connectivity and the asymmetrical nature of the hacking threat present a unique set of challenges that the current regulatory framework simply cannot address. Moreover, in the future, significant data breaches could be investigated similar to how the FAA currently addresses plane crashes: an independent investigator, empowered to release both public and classified findings, with close cooperation from the industry. This model could help realign incentives for sharing vulnerability information, if, as is the case with airlines and plane manufacturers, information is made equally available to all stakeholders.

What We Tend To Think



What Actually Happened



VULNERABILITY SHARING AND PRESENT BIAS

The classic prisoner's dilemma example: Two prisoners held in separate cells; each is interrogated separately regarding a crime they are accused of committing together. Each prisoner can either betray the other by ratting out their accomplice, or can cooperate by remaining silent. If both end up betraying the other, they're sent to prison for, say, two years. If one rats out the other but the second prisoner stays silent, the one who speaks (betrays) gets off free while the silent accomplice is sent to prison for three years. If they both stay silent, then they each receive a one year sentence. Because betrayal provides each prisoner with the best outcome, they are incentivized to betray one another. This leads to neither prisoner cooperating with the other.

Vulnerability sharing represents a similar (albeit less extreme) prisoner's dilemma. Enterprises need to determine whether the costs of sharing today (e.g. reputational risks) are outweighed by the future benefits of sharing (e.g. stronger security overall). The result is similar to that of the prisoner's dilemma: each organization is incentivized not to share, but to take the information from the other firms who do. In the end, none share, and all are worse off because of it.

However, unlike the classic prisoner's dilemma, vulnerability sharing presents two twists. The first is that the game is not one-off, but is a repeated exercise—for instance, CapitalCorp might choose to share their information one year, but not the next. Second, the costs and benefits do not accrue in the same period—the costs of sharing are immediate (e.g. reputational risk), but the benefits (e.g. a flow of vulnerability information) come over time.

Many argue that mutual collaboration is reachable so long as the firms focus on those future benefits as opposed to the immediate costs.³³ However, behavioral science can help explain why organizations may fail to be sufficiently patient when it comes to future benefits: **PRESENT BIAS**, or the tendency to let immediate costs outweigh far-off, long-term benefits. Anyone who has ever struggled to delay gratification will recognize this tendency: it can be incredibly difficult to defer an immediate reward, even when that means foregoing something much better in the future.

To improve cooperation, institutions interested in increasing vulnerability sharing should focus on interventions that reduce organizations' present bias. This can be done by reducing the immediate costs, making the future benefits more vivid, pre-committing organizations to long-term goals and blueprints for disclosure, or by bringing the loss of future benefits into the present.

“Who are they?” asked Rebecca.

“Amy Miller is the ExchangeWatch CEO, and she’s been cooperative with us to date, so there’s that. She’ll be joined by the firm’s IT administrator, a guy named Fred Altshuler. I want to see how far they’ve come in their internal investigation, and I want you to assess what conditions led to this whole fiasco.”

The two of them huddled around the phone while Kepler dialed the number. An assistant picked up and connected them to Amy and Fred who were already waiting on the call.

“As you can imagine, we’re still picking up the pieces,” said Amy. “I’m not sure if Fred has slept this week.” Rebecca and Kepler could hear Fred’s nervous chuckle through the phone.

Kepler asked Amy and Fred to provide an update, and, punctuated by nervous ticks, Fred gave a recap of what he had discovered to date. He told Rebecca and Kepler that identifying Peter Frank was the first and easiest step in the process. A quick examination of the outgoing press release logs showed that the release in question had been sent from Peter’s account, but not from Peter’s work computer. After Peter had denied knowing anything about the press release, Fred pulled an audit of log-ins to Peter’s account to piece together when the account had been compromised.

“Uh, it was clear that at least one person, p-p-potentially more, had been trying to break into his account for at least a few days,” said Fred. He explained that he had found a pattern of wrong password attempts spanning a two-day period from a set of different IP addresses. Fred told them that the existence of different IP addresses could have indicated that a set of people were trying to break in simultaneously, but he believed that it was actually someone using proxy servers to hide their real IP. The “smoking gun,” as Fred put it, was that the attempts had been spaced out relatively regularly over that two-day period. “Morning, noon, mid-afternoon and night, two days in a row, eight log-ins total,” said Fred. “Uh, it was as if someone had tried to break in during each of their meals. Had they not gotten in on the eighth attempt, I might have been notified about it.”

“What do you mean?” Kepler asked.

“Uh, I get n-n-notifications almost every day about someone failing to log in with the right password eight times in a row,” Fred said. “These sorts of trigger warnings are relatively normal for systems such as ours. People tend to forget their passwords pretty regularly. Ideally, it’s supposed to help us identify when someone’s

trying to break in. That said, if you space out your attempts like that, and the person who actually owns the account signs in between attempts, uh, I wouldn't ever get a warning about it. Whoever broke in knew they would be much less likely to be noticed if they were patient."

"You said that you get these trigger warnings regularly? How regularly?" Rebecca asked.

There was a short silence as Fred thought about it. "I don't know, maybe a few times per week? Maybe more? Pretty regularly."

"How frequently do you look into these to see if there's actually a security breach?"

"Rarely, if ever. I think, before this incident, the only I time I looked into them was when we first set up the system. I got a bunch right off the bat. But I, uh, quickly realized that it was just employees forgetting their new p-p-p-passwords and not a bad actor, so I stopped looking. Now I just delete the notifications from my inbox."

"Fred, when we spoke with Peter Frank, he told us that he didn't usually service the CapitalCorp account, but had access permissions to do so. Can you explain how permissions might have been given out in that case?"

"Uh, sure. Managers can set access permissions within the system, which would either allow or disallow users to send out communications on behalf of our clients. The idea is to prevent people from working on accounts they don't own. Sometimes, as was the case with Peter Frank, people expand access permissions for a particular user because the person who ordinarily services an account is out of the office, say on vacation for instance."

"As I understand it, these sorts of permissions should be rescinded as soon as the usual person comes back to work?"

"I mean, ideally. But frankly, it's hard for me to keep track of these things. There are probably some folks in the system that have more expansive permissions than they should, but I'm not sure." Fred explained that he spent a good majority of his days helping his fellow colleagues troubleshoot problems on their computers, setting up software, and managing the firm's external IT providers; the day-to-day grind gets in the way of monitoring people's permissions.

"Fred wears a lot of hats here, and we're very grateful for that," said Amy.

"You must have managed access permissions in the past, though?" Rebecca asked Fred.



ACCESS CONTROLS AND THE CONTEXT OF SCARCITY

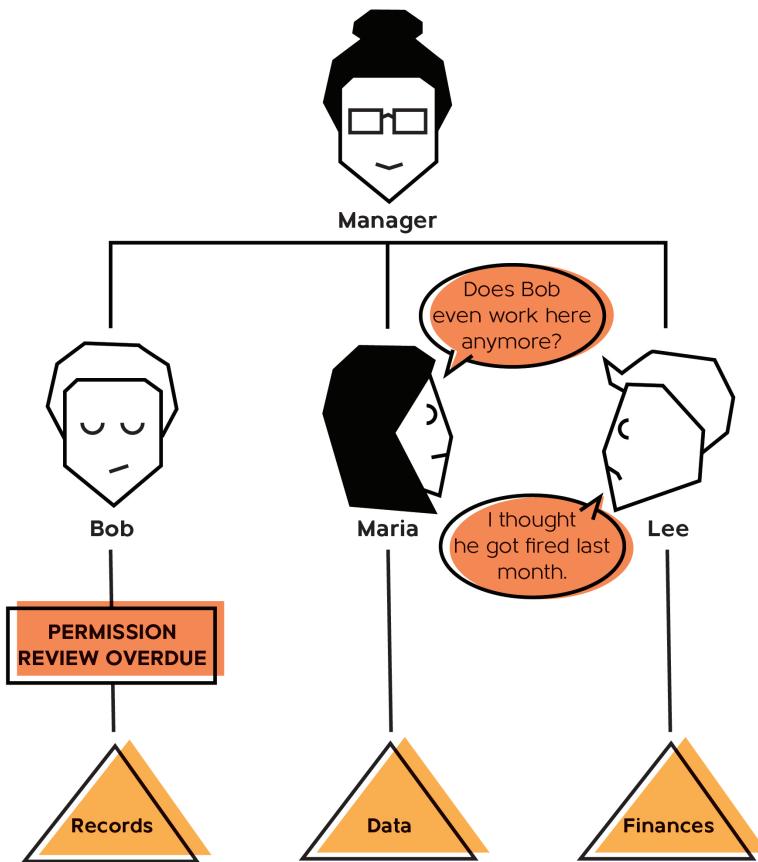
Ensuring that the right users—and only the right users—can access sensitive data can be a difficult task. In many organizations, even though the IT director oversees such permissions, the responsibility of attesting to them often belongs to managers.³⁴ After a breach, access controls are one of the last lines of defense between an attacker and the rest of the computer system and the information therein.

Behavioral science has shown that the **CONTEXT OF SCARCITY**, or lacking an essential resource (e.g. time, food, money), can place an undue psychological burden on people. People tend to “tunnel” in response to scarcity, focusing on tasks and activities that are most pressing, which in turn crowds out just about everything else.³⁵

Like Fred and the managers at ExchangeWatch, most people have a lot on their plates. Managers also have a host of other responsibilities that focus less explicitly on security (for example, releasing a product on time, or getting a press release out the door). This time scarcity causes people to direct their attention and tunnel on urgent tasks that likely have nothing to do with access control management. The result? It becomes a whole lot easier to for the day-to-day grind of managing controls to be continually postponed or neglected for extended periods of time.

One way to mitigate these concerns might be to automate parts of the review process by leveraging defaults. For example, permissions could automatically expire after a certain amount of time if not manually renewed. Alternatively, instead of making the regular review and approval of access controls the responsibility of the already taxed manager, this could instead be asked of each employee who likely has a better sense of what permissions they actually need on a regular basis. Either of these solutions might have afforded Fred a safety net in tumultuous times.





"Uh, yes, I did. I still do, but sometimes it's hard to know when p-p-permissions should be changed, you know? We have a policy that access permissions should be reviewed every 180 days, but in reality, that doesn't usually happen. Managers are supposed to be the ones responsible for taking care of that or letting me know who should or should not have permissions, but for most, it's a lot easier for everyone involved to just keep the expanded permissions. The managers don't know if and when someone might need them again, and it's less taxing for them to do nothing than going into the system to modify someone's access."

"Would you ever just rescind the permissions after 180 days if you didn't hear back from a manager?"

"Nope," Fred said. "I can't tell you how many times I've gotten a complaint from someone because they lost their ability to do some *critical* thing right after I removed their permissions. We tried to automate that at one point, but the policy

doesn't always line up well with actual business needs, and I'd rather not get in the way. Being in IT can be a pretty thankless job, you know."

Kepler proceeded to ask Fred a few additional questions about other insights his internal investigation had yielded. However, to Kepler's dismay, Fred couldn't construct much more with the available data. From ExchangeWatch's perspective, the case went cold. Rebecca, however, had more concerns.

"Amy," Rebecca said. "Would you mind answering a few questions for me?"

"Certainly," said Amy.

In Rebecca's experience, the best information security programs were much like an arms race—hyper-diligent, and always one step ahead of the enemy. However, most of the time, organizations that didn't have the capital, personnel or regulatory requirements were satisfied in building walls out of mud and concrete. A half-decent company might even go around looking for cracks once in a while.

"Do you know when the organization first deployed the current IT security infrastructure?" Rebecca said.

"We did an overhaul of the organization's content management system about three years back," said Amy, "and at the time, reevaluated our security infrastructure. In fact, Fred was the one who helped us go through the provisioning process."

"How did you decide what kind of security infrastructure and policies to put in place?"

"Fred wanted to make sure that we were in compliance with the NIST framework, so we built a system that would put us in compliance."

"I made sure it checked all the boxes," said Fred.

"How successful do you think that build-out was?"

"Up until a few weeks ago it was doing what it was supposed to do," said Amy.

"Had you reevaluated the system in the three years it has been up and running?"

"Run a formal audit? No, we haven't. Not yet, anyway," said Fred.

"IT infrastructure comes up during our quarterly financial planning meeting, but we've never found a need to make further investments since it was first built," Amy said.

"How did you come to that conclusion?" Rebecca asked.

"If the system wasn't working, then we'd consider what needed to be done to fix it."

"Like now?"

Compliance Checklist for Security Stew

- ✓ CARROTS
- ✓ ONIONS
- ✓ CABBAGE
- ✓ POTATOES
- ✓ RICE



CONGRUENCE BIAS AND INVESTMENT

People make complicated decisions every day—we struggle to determine how much to save for retirement, the right 401(k) allocation, or how best to pay for our children's schooling. These decisions demand that we acquire, process and act on complex information. Sometimes we intuitively resort to applying **HEURISTICS** or mental shortcuts or decision aids that allow us to make "reasonably good" decisions without taking into account all of the available information about a problem or decision.³⁶

Faced with time constraints and imperfect information, enterprises often employ a heuristic akin to: "If I'm compliant with security standards, then I'm secure." Adhering to standards is never a bad idea, but simply being compliant with a security standard doesn't guarantee security. Where risks are dynamic, and costs are high, decision makers should be wary of using shortcuts and ensure they are doing everything possible to minimize risks.

Additionally, enterprises under weigh the likelihood of a breach simply because they haven't had one in the past. This mental shortcut is called the **CONGRUENCE HEURISTIC** and occurs when someone tests a hypothesis by looking only for confirmatory information while simultaneously ignoring alternative hypotheses.³⁷ This concept has been talked about for centuries by philosophers like David Hume³⁸ and has more recently been applied by Nassim Nicolas Taleb, who presents an illustrative, if perhaps jarring, example:

"Consider a turkey that is fed every day. Every single feeding will firm up the bird's belief that it is the general rule of life to be fed every day. [...] On the afternoon of the Wednesday before Thanksgiving, something unexpected will happen to the turkey. It will incur a revision of belief."³⁹

To avoid making the same error as the turkey, organizations should use or develop tools to help guide investment decisions that prompt decision makers to consider alternative hypotheses. By making those alternatives salient, decision makers may be less inclined to use biased evidence and instead make better investment decisions.

100 and 1 days in the Life of a Thanksgiving Turkey



"Yes," said Amy, "now that all this has happened, we'll have to reevaluate the system."

"Not to put too fine a point on it, but I just want to make sure I fully understand. You're saying that you took an 'if it ain't broke, don't fix it' approach to this? So, so long as there had not been an apparent security breach, then you saw no reason to make any improvements?"

Amy's tone stiffened. "Ms. Saxon, we've given you all the information we can give you, and I have a team I need to get back to. Can we finish this up please?"

"Yes, that's fine," said Kepler. "But, before I let you go I have to ask: is there anything you can tell me about CapitalCorp?"

"What do you mean?"

"I've been trying to talk with someone on their side ever since the day the press release came out, but they've been as responsive as a rock."

"Then they probably don't have anything to share."

"Somehow, I doubt that," said Kepler. "You haven't heard anything?"

"Agent, frankly, I don't have time to concern myself with CapitalCorp's problems, I have enough of my own." And with that, Amy said she had other business to attend to and promptly ended the call.

Kepler stared ahead for a moment before speaking. "I'll let you know if I can get someone from CapitalCorp to open up to us, but I think this might be the end of the line," he said. "I look forward to reading your report when it's ready."

Kepler walked Rebecca back out of the office. Waiting for the elevator, she glanced over at him. She recognized in Kepler someone who couldn't accept a dead end. She knew he was the sort of person who would meticulously pick pebbles out of the treads of someone's shoes if he thought it would give him a lead, and she hoped that he'd keep at it.

"I'm sorry there wasn't more to go on," Rebecca said.

"It happens sometimes," said Kepler.

Rebecca nodded. "Well, let me know if there's anything else I can do."

Kepler smiled. "You think you might be able to get your daughter talking?"

"I can't tell if you're serious."

"Half serious."

"Doubtful, but even if I could, I don't know if I would. It just doesn't seem right."

"I understand," he said.

The elevator doors opened with a ding. Kepler thanked Rebecca for her help, and she stepped into the elevator car. She watched Kepler disappear behind the closing doors, before beginning the descent back to the street.

CHAPTER 6

GONE PHISHING

D33PTH0UGH1 EXITED THE SUBWAY PLATFORM at 59th St. and Lexington Ave. with the crowd of morning commuters. Her stomach ached with hunger, and her head throbbed from caffeine withdrawal. She was desperate to find something to eat but was painfully aware that there was little time to scour her surroundings for food. She started walking south down Lexington; on the first block were no places to grab a quick bite, but just down 57th St., she spotted a small coffee shop and headed toward it.

She wasn't close enough to smell the roasting coffee or freshly baked pastries, but her imagination was working overtime, and she was certain that her sense of smell was just powerful enough to take in the wafting aromas of the cafe. But, before she halved the distance between the end of the block and the shop, her burner phone rang. Coffee and pastries would have to wait.

D33pTh0ugh1 didn't know who the caller was, but she knew why they were calling. She answered the phone with the bubbliest voice she could muster, "Hello, this is Julia Short, corporate communications. How may I help you?"

The man on the other line introduced himself as a reporter for a regional newspaper and asked to confirm the validity of a press release that had just crossed his desk. Continuing to play the part, D33pTh0ugh1 confirmed that the press release had in fact been issued by CapitalCorp that morning, but that the company was not going to be able to comment.

“We’re publishing the story in the next twenty minutes, are you sure there’s nothing you want to offer up?” asked the reporter.

D33pTh0ught1 stood outside the coffee shop, watching with longing as people entered. “We appreciate your call, but as I said, we’re not making any additional public statements at this time, which includes Mr. James Robinson. We’re currently working through the details and will be making a public statement shortly.”

“We’ll run this story without comment from CapitalCorp then.”

“We understand. Thank you,” she said and hung up the phone.

Seconds later another call came through, this time from a major newspaper, and after that another call from a TV station in Boston. D33pTh0ugh1 continued to hold court outside the coffee shop for another 30 minutes, answering calls and giving the same canned response over and over again. By the time she finished, she had spoken with 23 different reporters from all over the country representing a constellation of publications, TV networks, news blogs, and radio programs.

The phone stopped ringing at 10 AM, just as planned, and by that point, the hunger pangs were nearly unbearable. But, before venturing into the cafe, D33pTh0ugh1 promptly dismantled the burner phone by removing the battery, snapping the SIM card in half, and throwing the whole mess down a nearby sewer grate.

She walked into the crowded coffee shop, and, standing in line, eyed what looked like the most delicious cheese Danish in all of New York City. She mused to herself about how smoothly the whole operation had gone. A day earlier she had logged into the CapitalCorp VoIP^{vii} admin portal without a hitch, and quietly set up call forwarding across all of the corporate communications lines so all incoming calls would be sent directly to her burner phone from 9:30 AM to 10 AM. The CapitalCorp communications team would only now be noticing that they were having an unusually light morning, but it was too late for them—it was a done deed.

“Thank you, David,” she muttered under her breath as she stepped up to the cash register.



DAVID NOTICED THAT HIS FACE was beginning to feel leathery. He had been sitting in the same beach chair for over an hour and hadn’t reapplied sunblock. His

^{vii} VoIP, or Voice over Internet Protocol is a technology that allows for the delivery of voice communication and phone services via the Internet and other Internet Protocol networks.

children were still in the water, playing “King of the Mountain” on an inflatable trampoline anchored fifty feet off the beach. Their splashes punctuated the rhythmic sound of the lapping sea as they successively tumbled from the inner tube into the water. David reached down beside the chair and blindly fumbled through the beach bag for his phone.

“They won’t be this age forever,” he mumbled as he snapped a photo of his children mid-flight, thrown from the trampoline back into the sea. “What do you think?” he asked his wife, holding the phone out across the sandy threshold between their two beach chairs.

She lazily turned her head to look and returned to sunbathing without comment.

“What did you think?” David asked again, determined to get a response.

“I couldn’t see it. The sun is too bright,” she said, motionless.

David sat up and brought the phone as close to his eyes as he could, shielding the sun with his hand like a visor. He inspected the shot for a second. “I think it looks good. I’m going to post it.”

“Awesome,” she said dryly.

David opened his photo-sharing app, slapped on a nostalgic filter, and posted the picture with the hashtags *#bahamas2016* and *#collinsfamilyvacation*. He then added image number eighty-five to social media in the album “Collins Family Vacation 2016.” He put the phone back into the bag and exchanged it for some sunblock, which he began reapplying to his face.

“You know, it’s already too late,” his wife said.

“What’s too late?”

“You’re sunburned. Putting sunblock on isn’t going to protect you at this point. You’re better off getting out of the sun.”

David’s phone chirped in the bag, and, soon after, chirped again.

“Can you take a look at that for me?” David asked.

“You want to see who liked your post?” she asked, looking over her sunglasses at David with a prodding smirk.

“No, those were emails. Can you take a look for me? I have sunblock on my hands.”

She pulled out the phone out from the beach bag. Two new email notifications showed on the screen, one from something called “Voicenet” and another from someone named Theresa. “Who’s Theresa?” she asked, tossing the phone into his lap.

“It’s my boss’s assistant,” David said with a hint of stress in his voice. He quickly finished applying the sunblock to his face and wiped his hands clean with his towel. He swiped open the phone, leaving a greasy smudge on the glass. The subject of Theresa’s email read, “[Urgent] Phone help.” David read the email.

Hello, David. I hope I’m not interrupting your vacation! Dana asked that I set up some complicated phone forwarding for her while she’s attending some meetings out of town next week. I went into her Voicenet account and tried to put everything in place, but I got a message telling me that you needed to approve the changes. Did you receive an email from Voicenet about it? If yes, can you approve the changes and let me know? Thanks in advance, you’re a lifesaver!

David went back to his inbox and saw the message from Voicenet unopened under Theresa’s email. The subject read “[noreply] action requested.” David opened the email to find a short note from Voicenet and a link to the admin portal. The notification read:

The user Dana Mattrick has attempted to make changes to settings on their account that require administrator permissions. To review and approve these changes, please sign into the Voicenet administrator portal.

David clicked on the link and ended up on the portal login page. He entered his username and password into the respective textboxes and clicked “Login.” The page took a minute to load and eventually sent him to an error page. David frowned. “The session must have timed out,” he said to himself.

David held his phone up higher and hit the back button to return to the login page, hoping to catch a stronger Wi-Fi signal from the resort this time. The login page loaded again, and after re-entering his credentials, and holding the phone back up again, he pressed the “login” button. This time, the admin portal home page loaded. But, scrolling through the page, it wasn’t immediately clear what he needed to do. There were no messages in his message center, and no popups providing him any information about the permissions. But, before he could make much of it, his phone chirped again, and another email arrived. It was from Theresa, and all it said was: **It looks like it’s working now! Thanks!** David replied with **No problem.** He closed his phone and tossed it back into the beach bag.

His wife turned to him. “What was that about?”

PHISHING FROM AUTHORITY

What would you do if your boss, or boss's assistant, asked you to complete a task ASAP? David did what we all do: he complied. Unfortunately, in this case, Theresa did not send the email, and David became the unwitting victim of a phishing attack.

Phishing is a type of attack where a bad actor tries to extract personal information from someone through email, chat, and even over the phone by posing as a trusted person or entity. Phishing remains one of the most frequently used attack techniques by bad actors, and there are many different strategies for extracting information effectively. In this scenario, D33pTh0ugh1 chose to masquerade as one of an authority figure. In fact, emails sent from authority figures, and especially those that include urgent requests, tend to work for the attacker.^{40,41,42} But why do people quickly, if not automatically, comply with requests from authority figures?

In his seminal book *Influence*, Robert Cialdini discusses how people can be influenced by those they perceive to have authority. Our deference to authority is likely conditioned, as we're all brought up to obey and defer to people who are in authority positions, (e.g. parents, teachers, etc.). The mechanism through which this **COMMAND AUTHORITY** functions is the perception of consequence—that if a request from someone in an authority position is disobeyed there might be a cost.⁴³ Authority, however, is not necessarily an objective characteristic. People tend to associate cues like role or job title, appearance, and assertiveness with authority. Additionally, people may overweight information conforms to their mental model of authority. Because we utilize these cues to approximate authority, those same cues can be used maliciously to provide the appearance of authority in a phishing attack.

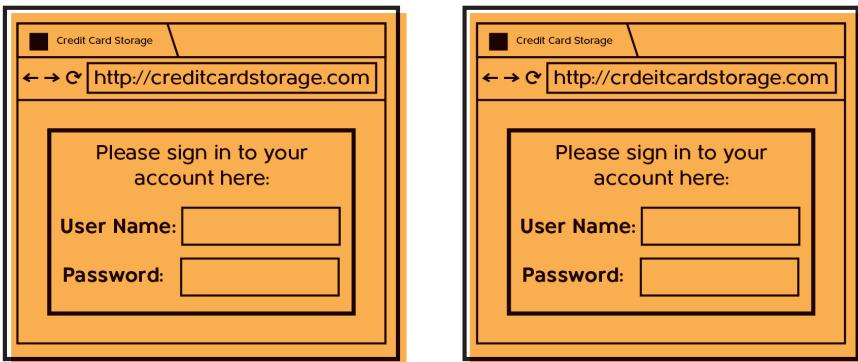
Phishing emails use corporate names and logos to build a façade of legitimacy. Information from a recognized authority can provide a valuable shortcut for deciding how to act in a given situation. One way organizations and service providers can help reduce the effectiveness of phishing attacks that use authority is to provide users with clear, up-front channels for how specific types of information will be collected or how notifications will be disseminated. These channels should not be easily spoofed by bad actors (e.g. take these communications offline, or only allow them within proprietary interfaces), but are still standard and accessible channels for end users.

“Nothing,” he said, “the boss needed a little help with her phone.” He pulled the sunblock back from the bag and finished reapplying it to the rest of his body before lying back down on the chair.

“I told you, you already have a burn,” said his wife.

“I know, I know,” he said, rolling over in the opposite direction.

Can you spot the difference?



AN EMAIL FROM DAVID POPPED UP in the `theresa42@mailserv.com` account that D33pTh0ugh1 had created, which read [No problem](#). D33pTh0ugh1's little phishing expedition had been successful, and she now had the information she needed to complete the attack.

A small bit of scanning through David's social media accounts showed that his vacation travel had been quite regimented for the past several years. August trips to the Bahamas had been a family staple ever since his youngest had entered elementary school as evidenced by a post in 2010 showing a picture of David on the beach holding the child up in the air. The caption on the photo that read: *First family trip to the Caribbean and Jessica can't stop talking about her upcoming first day of school! Is this my daughter?!!* After that year he had posted photo albums for each consecutive trip that they had taken over the six-year interim. Phishing David when he was



PRIMED TO SEE WHAT THEY WANT YOU TO SEE

Even though David clicked on a link that contained a typo, why wouldn't he recognize a spoof of a frequently visited page like his login screen? Surely, David would quickly notice a difference if D33pTh0ugh1 couldn't replicate the browser experience accurately, right? Not necessarily.

Research into site spoofing has shown that people often fall for well-spoofed pages because they tend to evaluate the legitimacy of websites based on the site's and the professionalism of the design, and not necessarily the page's URL.⁴⁴ What people look for when evaluating a product or an experience are **SALIENT CUES** (e.g. familiar visual interface, professional design, etc.), which may or may not provide valid information about the actual characteristics the user is trying to assess (e.g. security).⁴⁵ Moreover, the salient cues users do look for may not be the ones that would provide them with insights about the relative security or insecurity of a web page.

Additionally, D33pTh0ugh1 told David that he needed to sign into a portal, which ensured that David would direct his attention to the details of the login interface, as opposed to other visual cues. This phenomenon is an extension of visual **PRIMING**—the idea that “what we have recently seen and attended to strongly influences how we allocate visual attention.”⁴⁶ In this case, David was primed to expect a familiar process (e.g. the login screen), which in turn made him less likely to pay attention to other details and to notice that he was handing his username and password to D33pTh0ugh1 on a silver platter.

To design around this problem, web developers and UX designers might build processes into browsers or email interfaces that redirect users' attention toward the “right” salient cues. For instance, before loading a link embedded in an email, the email client might prompt the user to confirm that the URL that they are traveling to is valid. An additional level of support for users who are less familiar with URLs would be to provide rules of thumb to help users better evaluate whether the URL is, in fact, safe.



least likely to be paying much attention to work seemed like a prudent strategy for D33pTh0ugh1, and what better time than during a family vacation on the beach?

D33pTh0ugh1 had to build a trap and lure David into it, which was not a simple task. Creating a convincing spoof website to capture login credentials required keen attention to detail. To be convincing, the user needed to see what they anticipated seeing, which meant ensuring the admin portal looked and felt exactly like the real one. The user interface, links, and other page attributes needed to be exact replicas, and the URL had to be familiar too. Because she couldn't use the exact URL, D33pTh0ugh1 decided that typosquatting^{viii} on the admin portal URL might work. The actual portal URL was portal.voicenet.com, but by taking the original URL and switching around the placement of the 'o' and 'r' in 'portal,' she could register a new website at protal.voicenet.com, a small enough change that David was unlikely to notice. But once he entered his credentials, where would he go? It would be nearly impossible to build a fully functioning spoof of the admin portal itself with all the essential details, so she needed to figure out some other diversion that wouldn't draw suspicion. After thinking about it for a bit, she decided that she could build an error page to make it look like the connection didn't go through, and embed a link back to the real login URL so David could try to log in again and do so successfully.

Sending the emails out to prompt David to log in was a little more complicated. Masquerading as someone else is not terribly difficult over email, but it often requires finding an open outgoing mail server, which, nowadays, were few and far between. Open SMTP servers were mostly a thing of the past, as contemporary mail server software closed the SMTPs by default.^{ix} However, it was still possible to sniff out occasional open SMTPs, and D33pTh0ugh1 knew a professional spammer in China through personal connections in the deep web who might be able to help. She got in contact with him, and they worked out a deal that he would let her know if one opened up during the period that David was on vacation, but that he couldn't make any promises about how long it would be open.

^{viii} Typosquatting is technique designed to direct users to fake websites by intentionally inserting a typographical error that often goes unnoticed, or is likely to be typed by accident. Here, D33pTh0ugh1 leveraged the insight that humans can generally read words that jumble the contents between the first and last letters.

^{ix} SMTP (Simple Mail Transfer Protocol) is the method used by email servers to send our emails. By Open SMTP server, the hacker is referring to open mail relay servers which are configured to allow anyone to send emails through them. In contrast, closed mail relay servers only send and receive emails from known users. In the early days of the Internet, SMTP servers were open relays, but today, most are closed to avoid exploitation from spammers and worms.



INSECURITY BY DEFAULT

David unwittingly made the attack a little bit easier with his social media habits. Posts and even entire photo albums of his family were visible to the public. Why didn't David switch his privacy settings? One reason is that users sometimes have incorrect mental models about the default level of security and privacy they may have when using a service like a social networking site⁴⁷ or an Internet-connected product. When incorrect, mental models about the security defaults can be especially problematic because defaults are very sticky.

To illustrate how defaults work, consider retirement savings. Policymakers and employers observed that they could increase retirement savings by changing the default. Originally, employees had to opt-in to their company's 401(k) plans, but relatively few people did so. By changing the default from opt-in to an opt-out, not only did enrollment rates in 401(k) increase significantly, but the default contribution rates had a strong impact on savings.⁴⁸

Defaults can be a powerful tool for swaying behavior both positively and negatively, and this is no less true when it comes to cybersecurity. One example of this is a recent distributed denial-of-service (DDoS) attack on the DNS provider Dyn, which caused massive outages and network congestion for many websites. The Dyn attack was executed using Mirai malware, which turned millions of Internet of things (IoT) devices (many of which were WI-FI enabled cameras) into a botnet that spanned the globe. Attackers recognized that many of the various IoT devices were still password protected with the default passwords that had been set by the manufacturer—they had never been reset by the users—making them easy to compromise.⁴⁹ Had the manufacturer automatically required users to reset the passwords as soon as the device was turned on or provided a random password for each separate device instead of a standardized default, this kind of event may have been avoided.

Default security settings are powerful because people are unlikely to change them. Organizations need to determine whether opt-in policies are reasonable when it comes to security, fully taking into account how people actually act. Instead, service providers and device manufacturers could make lower levels of security and privacy an opt-out decision from the beginning. Or, if opting out isn't feasible, service providers and device manufacturers could force consumers to make a considered decision about their security preferences during their first-time experiences through thoughtful UX design.



“These things close up almost as soon as they open,” he said, “so you’re going to have to take the opportunity when it comes.”

She was all right with that and set to work crafting the emails. The one from Voicenet needed to look automatic but also provide a sense of urgency. To focus David’s attention on following through on the Voicenet email instructions, and reduce the likelihood that he’d scrutinize the email too much, she decided to craft another email to contextualize the request. D33pTh0ugh1 decided that sending it from CapitalCorp’s CISO herself might seem a bit odd—what executives make those sorts of requests for themselves?—so instead, she pretended to be the CISO’s executive assistant, Theresa, which turned out to be an effective decision.

In the middle of August, when David was on vacation, the message came through from her Chinese contact that there was an open SMTP that she’d be able to co-op for her attack, and she immediately set to work.

“Have at it,” he said, and so she did.



THE BARISTA RETURNED TO the cash register with a cup of coffee in a to-go cup and a cheese Danish in a small wax paper bag.

“Best cheese Danish in all the city,” said the barista, “I promise.”

D33pTh0ugh1 rummaged through her purse and happily handed over the eight dollars and change she owed, leaving some extra in the tip cup on her way out. Once back on the street she took a small sip of coffee, which burned as it hit her tongue, and then a bite of the Danish to try to soothe the already numb taste buds. Despite the coffee being too hot, both it and the Danish were delicious, even more so than she had anticipated—*the barista was right*, she thought.

For a moment, looking out at the taxi cabs and morning commuters, she felt calm. All of her well-laid plans had unfolded, and now all she had to do was to wait and watch to see where the whole thing would land. In a sense, despite the work that she put in, it wasn’t that hard. The systems that she compromised, the passwords she collected, the trickery she had played on the reporters, all of her successes came down to the fact that people are predictable. But, there was a poignancy in it. All of these people, going about their happy little lives, had no idea how close they were to making a misstep and becoming a victim; they had no idea how, for

a person like D33pTh0ugh1, they were all like wounded animals on the savannah, completely unaware of their limp.

“If they were more like Spock,” she thought to herself, “then I’d be out of the job.” But part of her wished that she could be out of the job.

She took another bite of her cheese Danish, and her personal phone rang in her purse. She took the phone out, looked at the number, and answered.

“We’re going to be starting soon,” said the voice on the other line.

“I’m heading in now,” D33pTh0ugh1 replied. “I should be there in 15 minutes.” She hung up, put the phone back in her bag, and began her walk downtown.

CHAPTER 7

THE WAGER

IT WAS 10:28 AM BY the time D33pTh0ugh1 completed the seventeen-block hike downtown. In the shadow of the CapitalCorp tower, she finished her last gulp of coffee and threw the remains of the Danish and the cup into the trashcan on the sidewalk corner. The tower was not the tallest in the city, but it loomed over the neighboring buildings in such a way that granted it an immenseness that bordered on intimidating. Looking up, D33pTh0ugh1 watched with a growing sense of vertigo as the building seemed to sway against its backdrop of deep blue and cotton ball clouds.

She took a deep breath, and, patting down her clothes, wrinkled from sitting for too long on the subway platform, she crossed the street and spun through the revolving door into the building's foyer. The entrance opened into a vast atrium wrapped in a lattice of steel and glass, which allowed natural light to filter into the room. D33pTh0ugh1 swiftly passed the building's security desk to get to the elevators, where a row of hip-high security barriers and a well-dressed guard with an earpiece stood protecting the elevator bay. She approached one of the barriers and began rummaging through her bag, which, after a few moments, caught the guard's attention.

"Miss," said the guard. "Is there something that I can help you with?"

"No," she said. "I think I just—Ah!"

D33pTh0ugh1 pulled a security card out of her bag, which she displayed to the guard with a smile. He returned to his post, and she pressed the card against the card reader. The barrier doors opened with a “swoosh,” and she stepped through making a beeline for the elevator. D33pTh0ugh1 scanned the security card again and keyed-in the forty-second floor.



“YOU MUST BE OUR HACKER,” said James, the CEO of CapitalCorp.

When the woman before him didn’t respond, he asked, “What do they call you? ‘Profound Idea’ or something like that?”

“Something like that.”

They were in the glass box of a conference room on the executive suite’s lofted second floor.

“You laid it on pretty thick, you know. You didn’t have to do that,” James said.

“What’s the fun in that? I wanted to make sure it was newsworthy.”

After a moment of tense silence, the CEO’s face broke into a smile. “Fair,” he said, throwing his feet up on the table. “Besides, I won the bet, so there’s that. Dana, where did you find this one?”

Dana, CapitalCorp’s CISO, was preparing the conference line interface for their call. “Downstairs,” she said without taking her eyes off her screen.

James cocked his head, scrutinizing the hacker sitting across from him more carefully. “She’s one of ours?”

Dana looked up. “This is Sarah Saxon,” she said. “She’s our best security analyst. I wouldn’t have trusted anyone else with this.”

“Well, Ms. Saxon,” James said with a smile. “I hope you do in fact have the Answer to the Greatest Question of Life, the Universe, and Everything. I look forward to reading your report once it’s ready.” He turned to Dana. “Who are we waiting for?”

“I just sent for Damien, and Amy is joining remotely.”

“On a secure line, I hope?”

Dana glared at him. She finished setting up the conference line, turned on the large television bolted to the room’s one concrete wall, and pressed the button on the interface to join the call. A dial tone sounded through the room’s speakers and after a moment, Amy, ExchangeWatch’s CEO, appeared as an oversized, disembodied head on the television screen.

"Hello, Amy," said James with a self-satisfied grin. "It looks like I'll be commanding that case of 2004 Screaming Eagle that's sitting in your basement."

"It's good to see you too, James," said Amy. "You know, you probably could have just bought yourself two cases for the price you paid for that hacker."

James's grin widened. "Didn't pay a dime. She was an internal resource."

Sarah gave a timid wave to the camera.

"I don't think we've been introduced," said Amy.

"This is Sarah," said James.

Amy scowled. "Well, Damien already called me. I'd say he's a little on edge."

"As he should be. I pay him to be the one on edge."

"Let me throw him a bone. Sarah, whose account got compromised?"

Sarah looked to Dana, unsure if she should respond.

"It's okay, you can tell her," said Dana. "The whole thing's done, and her team will find out soon enough anyway."

Sarah turned back to the screen. "It was a guy named Peter Frank."

"Poor kid. I figured it would end up being one of the older employees." Amy picked up her cell phone, started typing, and then put the phone down again. "Are we waiting for anyone else?"

"Just Damien," said Dana. "He should be on his way up now."

Amy's phone began to ring over the conference line. "Speak of the devil. Give me a second," said Amy, muting her line. She had a short exchange over the phone before returning to the conference. "It was Damien. He doesn't have a clue what's going on."

As Amy spoke, Damien appeared on the other side of the executive suite. He crossed the couple hundred feet between the elevator and the second-floor stair and ascended to the conference room. Damien looked borderline sick. As he pulled open the glass door to the conference room, he scanned the faces of everyone present, looking at them all as if they were Martians. James gave him a warm welcome and told him to take a seat. Damien didn't move.

"What is this?" asked Damien.

"Just take a seat, we'll get to it," said James.

"Why is Amy on the screen, and why are you smiling at me, James?"

"Sit down. We'll talk it through."

"Who is she?" Damien asked, pointing a hostile finger at Sarah.

“Just sit.”

Damien gave an unsure glance around the room, pulled out one of the chairs and sat down. He took a small pad of paper and pen out of his jacket pocket and began setting them down on the table when Dana stopped him.

“No notes,” she said. “We’re just going to talk.”

Damien’s eyebrows knitted together and the corner of his mouth began to twitch. “Can someone please let me know what’s going on here?” he demanded.

James stood up from the table and walked toward the far end of the conference room where, through the floor-to-ceiling window, he could look out over the street below. For a moment, he watched the cars moving like electrons along the circuitry of a silicon chip.

“Damien,” he began, “first off, let me apologize for making your life harder than it needed to be. I can confirm with absolute certainty that the press release was completely fabricated.”

“I don’t understand,” said Damien. “The release is fake? You apologize? Why?” He was shaking.

James turned around to face Damien. “Because we sent it out.”

Damien just stared at James, mulling over the words in his head, half grasping at their meaning. Then, in a split second, whatever flush was left on Damien’s face completely washed out.

“What did you say?” asked Damien. “You’re not supposed to be sabotaging the company... you’re supposed to be leading it!”

“Then we’re on the same page. Sometimes you have to try to break things to figure out how resilient they are.”

“That’s absurd!”

“Actually,” said Dana, “it’s not. While this all came about in a ridiculous way, it ended up being the right thing to do. I wholeheartedly agree with James.”

“What do you mean, ‘a ridiculous way?’” Damien asked.

Amy piped up from the television screen. “She means it all happened because James can’t help but make stupid bets.”

A smile crept across James’s face. “But it was still a good idea.”

“We’re not out of the woods yet,” said Amy, rolling her eyes.

James explained that it all started at a dinner party both he and Amy had attended almost six months earlier. One of the other guests had recounted a story about

an executive friend who had to resign following a breach that occurred through one of the company's vendors. The executive's IT team had been warning about needing to change some of the software the vendor used, but the company never made the investment. Six months later, nearly a quarter million credit card numbers were stolen through the vendor's software access point. James and Amy were chatting about the story, and James had made an offhand joke about needing to vet Amy's security systems before he could continue to use ExchangeWatch as their primary press release service. Amy said the security was fine, and James made a wager that a good hacker could probably bypass the security system without breaking a single piece of hardware or line of code. In a move that surprised even James, Amy agreed to the bet.

"What were you going to give up if you lost?" asked Sarah.

"He said I could use his home in the Hamptons whenever I wanted to for the year," said Amy. "It was hard to resist."

James scoffed. "I wasn't going to lose."

"James came to me the next day to see if I'd help out," said Dana. "It didn't seem like a terrible idea. Sometimes people think about building security systems like castles, but that doesn't take into account all the risks that are out there. Despite James's hubris, it was an attractive proposition. It would give us a chance to build a model that allowed us to test our vendor's security while addressing another question altogether: how much of our collective risk comes down to people acting in insecure ways as opposed to failures in the software and hardware itself?"

"But why go all the way! We didn't actually need to send out a press release to check that, did we?"

"No," said Sarah. "But it allowed both organizations to test out their business continuity plans. It was the only way to figure out if our respective teams would take the necessary actions. You can't test that in a simulation."

"That's preposterous," said Damien. "And wait, who the hell are *you*?"

"I'm Deep Thought," said Sarah. "I'm the one who did all of this."

"You're the hacker?"

"She's the best security analyst we've got," said James.

"She works here?" Damien was nearly yelling now. "Great. Just great."

"Don't worry about that. Just do your job, and clean this up."

"You've got to give me a rundown of everything that happened," Damien said



MENTAL MODELS AND THINKING ABOUT SECURITY

Dana's analogy comparing security systems to castles exemplifies a **MENTAL MODEL** (also known as a *schema*), a psychological representation of a situation. Mental models mediate memory, perception, inference and evaluation and are often used to help make sense of scenarios that lack perfect information, such as designing a comprehensive cybersecurity strategy.

As a toy example to help illustrate the concept, suppose you held the mental model that "hammers are objects used to hit things" and were told to evaluate a hammer's effectiveness for taking a nail out of a wall. If you held such a "mental model" you might be more likely to give a negative assessment of the hammer than if you held the mental model that "hammers are tools."

Dana's example model of cyber defense as a castle evokes similar shortcomings. By thinking about building a cybersecurity infrastructure using the model of a castle, you may end up paying a lot of attention to boundary devices like firewalls, but not think about other potential vulnerabilities, such as the risks associated with users accessing critical information from computers outside of the office.

Everyone has mental models, and some mental models are more useful than others. In the best cases, they help us make sense of complicated, messy situations and facilitate action. However, using the right model for the right context is critical for making good decisions, and using the wrong mental model might make people less likely to protect themselves, or protect themselves in the right ways.⁵⁰ One way we might be able to better support both decision-makers and systems engineers would be to develop tools that require them to consider ways in which their existent models may break down or be incomplete, and find alternative models that have better explanatory power.



to Sarah. “I need the whole story before I can figure out what I’m going to have to do next.”

“I think the less you know, the better,” Dana said.

“No, no, no. That’s not how I operate,” said Damien. “Sarah, can you just tell me everything from start to finish? I think I’ll feel better once I understand what’s actually happened here.”

James gave Sarah a nod, and she began telling the story from the beginning. She told them about how she found Peter Frank’s social media credentials from a dump on the deep web, and how his passwords had been so bad that she had been able to reconstruct his ExchangeWatch login information without breaking a sweat. Sarah explained how she had masqueraded as Dana’s assistant, and how she had spoofed the VoIP login portal to capture David’s username and password. Sarah told them, much to Damien’s displeasure, how she had pretended to be the Capital-Corp’s communications department earlier that morning. She described how each attack leveraged different predictable ways in which people act and interact with the Internet, and how the best hackers often begin their most devastating attacks by simply tricking a human.

“We might put a lot of stock in the systems and technology that we build,” she said, closing the story, “but in the end, these systems are only as secure as their weakest user.”

Before they left, James made clear that nothing they had discussed would leave that room, and that all outgoing communications would be Damien’s sole responsibility.

“We’ll figure out a better way to do this next time,” said Dana. “We can’t keep on playing with fire like this.”

“You’re telling me,” said Damien. “Just keep me in the loop next time, okay?”

Damien left the conference room in a huff. Amy signed off the conference line, and James returned to his office. Dana and Sarah walked back to the elevators and descended back to the security department floor, ready to test the constitution and resilience of their teams in the wake of such an unprecedented “intrusion.”

CHAPTER 8

THE LONG WAY HOME

SARAH'S TRAIN ARRIVED AT the Katonah station around 7:00 PM. She walked off the platform to the small parking area across from the row of colonial-style storefronts where her mother was already waiting.

"How was the ride up?" Rebecca asked as Sarah took a seat in the car.

"Easy," Sarah said.

The two drove Northwest, across the Muscoot Reservoir and up RT 35 to Amawalk and then on to Rebecca's home, picking up some chicken, a bottle of wine, and a few other ingredients for dinner along the way.

"Guess what happened to me today," said Rebecca as she drove down the quiet country lane.

"Your Ph.D. students put you in a Skinner Box?" Sarah joked, peering out the car window into the distance.

"No, silly." Rebecca paused for a moment. "I probably shouldn't be telling you this, but I was down at the FBI office today."

"Cool. What were you doing there?"

"Well, they called me in to do a little forensic work around the ExchangeWatch hack."

Sarah could feel the blood rushing to her face. While the scenery continued to pass, Sarah was blind to it all, lost in her thoughts, wondering in quiet panic

whether her mother or the FBI for that matter had figured something out that could implicate her. For a second, the image of being dragged out of her apartment by a procession of badged strangers flooded her consciousness. She became aware of the choking feeling of her heart in her throat.

“Oh?” she said as coolly as she could, “Any juicy facts I should know about?”

“Probably nothing you don’t already know.”

Sarah waited for her mother to go on, but Rebecca didn’t say more. The car continued to ramble on passing small residential enclaves, tucked away among the pines and oaks, and the occasional farmhouse along the flat, open land. Eventually, they turned into Rebecca’s neighborhood, and then arrived at her home. Rebecca and Sarah carried the luggage and groceries into the house.

While Sarah helped her mother chop the vegetables and prepare the chicken for the oven, she inquired whether the FBI had been in touch with anyone at CapitalCorp.

“Well,” said Rebecca, “I had meant to ask you about that. The FBI told me that they haven’t been able to get anything out of CapitalCorp at all.”

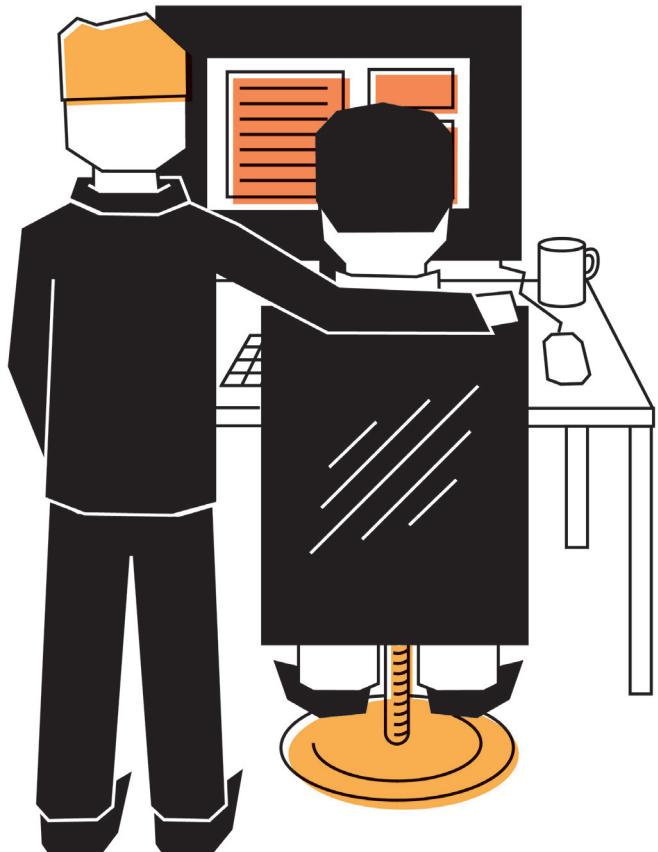
Sarah immediately felt a small sense of relief. “Yeah, that makes sense. We’re kind of uninviting when it comes to people looking into our security systems. We have a lot of sensitive information that we would like to keep under wraps.”

“You don’t happen to know anything, do you? CapitalCorp hasn’t been trying to run its own investigation or anything like that?”

“Nope,” said Sarah. “No one broke into our systems, so there wasn’t much forensic work to do anyway.”

Rebecca nodded in bittersweet acceptance. As they finished preparing dinner, she resigned herself to the idea that she would never know. This was not her first unsolved case, and she knew that it was not likely to be her last. Life would go on, and she would eventually forget about the nagging curiosity as she found something else to preoccupy her intellect.

The two of them spent the remaining hours of that evening drinking wine and enjoying each other’s company, forgetting the experiences of the day, and focusing instead on each other. They feasted, and finally trekked up the stairs to sleep just as the day slipped away, and the next one was upon them.



IT WAS ALREADY WELL past dark by the time Kepler began packing up his bags to head home. The investigation had been going on for nearly a month at this point, but there had been no additional leads, and CapitalCorp continued to give Kepler the cold shoulder. He was beginning to feel like it was about time to give up on the whole investigation and close the case cold. It was an unsettling feeling for him. While he had more than a few cold cases in the past, he never felt good about walking away from an unfinished investigation. But as he put on his jacket and prepared to leave the office, he became resigned to the idea that he would need to move on.

The office was nearly vacant by this time in the evening, save for a couple of analysts at desks and investigators in their offices, preparing to work far longer into the night. He figured that he'd treat himself to a long walk home, back to his apartment in Midtown East. The winter chill hadn't begun to fully set in, and the city had been going through an unusual bout of pleasant late fall weather.

But as he made his way through the stand of cubicles in the main office, he heard someone calling his name. He turned to see one of the analysts on the case lit up by their computer screen, beckoning him to come over.

"I've found something I think you should see," the analyst said.

Kepler walked over to the analyst, who was pointing at the screen with an outstretched finger.

"There," he said. "I think we have something."

Kepler peered into the data-filled screen. Above the young man's finger was some sort of name: "D33pTh0ugh1."

"What do you think it is?" asked the analyst.

"It's something," said Kepler. "How long do you plan on staying tonight?"

"I was going to head out in an hour or so."

"Okay, let's get back on this in the morning. Good work, kid."

Kepler turned around and left through the office doors, descending to the street level on the elevator. He walked outside through the building's front entrance and looked up at the clear night sky. He took a deep breath and felt the cool air fill his nose and lungs, and he exhaled, letting all of the stress and uncertainty leave his body. Knowing that tomorrow would be a new day, he turned north on Broadway, and began his long march uptown.

-
- ¹ Leventhal, H., Singer, R., & Jones, S. (1965). Effects of fear and specificity of recommendation upon attitudes and behavior. *Journal of Personality and Social Psychology*, 2, 20-29.
- ² Leventhal, H. (1970). Findings and theory in the study of fear communications. *Advances in Experimental Social Psychology*, 119-186.
- ³ Legezo, D. (November 24, 2016) Research on Unsecured Wi-Fi Networks Around the World. Secure List. Accessed from: <https://securelist.com/blog/research/76733/research-on-unsecured-wi-fi-networks-across-the-world/> on January 6, 2017
- ⁴ Bruce Schneier (2000) "Semantic Attacks: The Third Wave of Network Attacks." Schneier on Security. Accessed from: <https://www.schneier.com/crypto-gram/archives/2000/1015.html#1>
- ⁵ Thaler, R. H. (2015). *Misbehaving: The making of behavioral economics*. WW Norton & Company.
- ⁶ Thaler, R. H. (2015). *Misbehaving: The making of behavioral economics*. WW Norton & Company.
- ⁷ Association of Internet Security Professionals (2014). "Illegal Streaming and Cybersecurity Risks: A dangerous status quo?". AIS Working Paper, Aug 2014
- ⁸ U.S. Department of State. "Foreign Per Diems by Location." Accessed from: https://aoprls.state.gov/web920/per_diem.asp on October 2nd, 2016
- ⁹ Slovic, P., Finucane, M. L., Peters, E., & Macgregor, D. G. (2007). The affect heuristic q, 177, 1333-1352. <http://doi.org/10.1016/j.ejor.2005.04.006>
- ¹⁰ Slovic, P., Finucane, M. L., Peters, E., & Macgregor, D. G. (2007). The affect heuristic q, 177, 1333-1352. <http://doi.org/10.1016/j.ejor.2005.04.006>
- ¹¹ Jenkins, J. L., Anderson, B. B., Vance, A., Kirwan, C. B., & Eargle, D. (2016). More harm than good? How messages that interrupt can make us vulnerable. *Information Systems Research*, 27(4), 880-896.
- ¹² Felt, A. P., Ainslie, A., Reeder, R. W., Consolvo, S., Thyagaraja, S., Bettes, A., ... & Grimes, J. (2015, April). Improving SSL warnings: Comprehension and adherence. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2893-2902). ACM.
- ¹³ Sunshine, J., Egelman, S., Almuhammedi, H., Atri, N., & Cranor, L. F. (2009, August). Crying Wolf: An Empirical Study of SSL Warning Effectiveness. In *USENIX security symposium* (pp. 399-416).
- ¹⁴ Felt, A. P., Ainslie, A., Reeder, R. W., Consolvo, S., Thyagaraja, S., Bettes, A., ... & Grimes, J. (2015, April). Improving SSL warnings: Comprehension and adherence. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2893-2902). ACM.
- ¹⁵ Thompson, R. F., & Spencer, W. A. (1966). Habituation: a model phenomenon for the study of neuronal substrates of behavior. *Psychological review*, 73(1), 16
- ¹⁶ Anderson, B. B., Kirwan, C. B., Jenkins, J. L., Eargle, D., Howard, S., & Vance, A. (2015, April). How polymorphic warnings reduce habituation in the brain: Insights from an fMRI study. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2883-2892). ACM.
- ¹⁷ Anderson, B. B., Vance, A., Jenkins, J. L., Kirwan, C. B., & Bjornn, D. (2017). It All Blurs Together: How the Effects of Habituation Generalize Across System Notifications and Security Warnings. In *Information Systems and Neuroscience* (pp. 43-49). Springer International Publishing.
- ¹⁸ Leonard, T. C. (2008). Richard H. Thaler, Cass R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness. *Constitutional Political Economy*, 19(4), 356-360.
- ¹⁹ Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York City, NY.
- ²⁰ Tversky, Amos; Kahneman, Daniel (1973), "Availability: A Heuristic for Judging Frequency and Probability", *Cognitive Psychology*, 5: 207-232
- ²¹ Kahneman, D. (1973). *Attention and effort* (p. 246). Englewood Cliffs, NJ: Prentice-Hall.
- ²² Kahneman, D.; Tversky, A. (1979). "Intuitive prediction: biases and corrective procedures". *TIMS Studies in Management Science*. 12: 313-327.

- ²³ Benartzi, S., & Lehrer, J. (2015). *The smarter screen: Surprising ways to influence and improve online behavior*. Portfolio.
- ²⁴ Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York City, NY.
- ²⁵ Komanduri, S., Shay, R., Kelley, P. G., Mazurek, M. L., Bauer, L., Christin, N., ... & Egelman, S. (2011, May). Of passwords and people: measuring the effect of password-composition policies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2595-2604). ACM.
- ²⁶ People will not switch passwords unless they are required to do so. See: Adams, A & Sasse, M.A. (1999). Users are not the Enemy. *Communications of the ACM*. 42(12), 40-46.
- ²⁷ Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of risk and uncertainty*, 1(1), 7-59.
- ²⁸ In one survey, at least half of all respondents reported writing down a password in one form or another. See: Adams, A & Sasse, M.A. (1999). Users are not the Enemy. *Communications of the ACM*. 42(12), 40-46.
- ²⁹ Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives*, 5(1), 193-206.
- ³⁰ Bettinger, E.P., Long, B.Y., Oreopoulos, P., & Sanbonmatsu, L. (2012). The Roles of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment. *The Quarterly Journal of Economics*. 127(3), 1205-1242.
- ³¹ Warkentin, M., & Willison, R. (2009). Behavioral and policy issues in information systems security: the insider threat. *European Journal of Information Systems*, 18(2), 101.
- ³² Kvochko, E. & Pant, R. (2017). Why Data Breaches Don't Hurt Stock Prices." Harvard Business Review.
- ³³ Fudenberg, D., & Maskin, E. (2017). The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica*, 54(3), 533-554.
- ³⁴ Oracle. (2010). *Identity and Access Management: Enabling Sarbanes-Oxley Compliance*. Retrieved from <http://www.oracle.com/us/products/middleware/identity-management/061145>
- ³⁵ Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why having too little means so much*. Macmillan.
- ³⁶ Tversky, A., & Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases. In *Utility, probability, and human decision making* (pp. 141-162). Springer Netherlands.
- ³⁷ Baron, J., Beattie, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: II. Congruence, information, and certainty. *Organizational Behavior and Human Decision Processes*, 42(1), 88-110.
- ³⁸ Hume, David (1748). *Philosophical Essays Concerning Human Understanding* (1 ed.). London: A. Millar.
- ³⁹ Taleb, N. N. (2007). *The Black Swan: The impact of the highly improbable* (Vol. 2). Random House.
- ⁴⁰ Ferguson, A. J. (2005). Fostering e-mail security awareness: The West Point carrouade. *Educate Quarterly*, 28(1), 54-57.
- ⁴¹ Workman, M. (2008). Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security. *Journal of the American Society for Information Science and Technology*, 59(4), 662-674.
- ⁴² Butavicius, M., Parsons, K., Pattinson, M., & McCormac, A. (2016). Breaching the Human Firewall: Social engineering in Phishing and Spear-Phishing Emails. *arXiv preprint arXiv:1606.00887*.
- ⁴³ Cialdini, R. B. (2007). *Influence: The psychology of persuasion*. New York: Collins.
- ⁴⁴ Dhamija, R., Tygar, J. D., & Hearst, M. (2006, April). Why phishing works. In *Proceedings of the SIGCHI Conference on Human Factors in computing systems* (pp. 581-590). ACM.
- ⁴⁵ Tom, G., Barnett, T., Lew, W., Selmants, J., (1987) "Cueing the consumer: The role of salient cues in consumer perception," *Journal of consumer marketing*, vol. 4 iss: 2, pp.23 - 27
- ⁴⁶ Kristjánsson, Á., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in

visual search tasks. *Attention, Perception, & Psychophysics*, 72(1), 5-18.

⁴⁷ Lin, J., Amini, S., Hong, J. I., Sadeh, N., Lindqvist, J., & Zhang, J. (2012, September). Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 501-510). ACM.

⁴⁸ Madrian, B. C., & Shea, D. F. (2001). The power of suggestion: Inertia in 401 (k) participation and savings behavior. *The Quarterly Journal of Economics*, 116(4), 1149-1187.

⁴⁹ Brian Krebs (October 21, 2016) Hacked Cameras, DVRs Powered Today's Massive Internet Outage. Krebs on Security. Accessed from <https://krebsonsecurity.com/2016/10/hacked-cameras-dvrs-powered-todays-massive-internet-outage/> on October 31, 2016

⁵⁰ Wash, R. (2010, July). Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security* (p. 11). ACM.

// APPENDIX

BEHAVIORAL CHALLENGES IN CYBERSECURITY

UPDATING



The Problem

As of 2015, as many as 85% of data breaches could have been prevented by installing available security patches.¹ Hackers can exploit software vulnerabilities by writing code, which is then packed into malware, to target the particular weakness. Nonetheless, the average consumer often drastically undervalues updating systems, despite experts advocating for its importance. In fact, a significant share of mobile users (73% of Android users² and 6% of iOS users³) are running outdated operating systems. The problem is likely to intensify as more devices join the Internet of Things (IoT), further increasing the need for the development, deployment, and installation of security patches.

Take the MySpace hack. In 2013, MySpace released a security patch but only applied it to new accounts. Consequently, accounts created before the new security protocol were less secure. Fast-forward three years, and a hacker, known as Peace, tried to sell 427 million MySpace usernames and passwords.⁴ This hack, which put the personal information of millions of users at risk, could have been prevented had MySpace applied the security update to all accounts.

Though updating may seem like a hassle, doing so is a critical preventative step for Internet safety.



Simple statement of the behavior we want to change

Many users do not update their software when updates are available. In some cases, users never install any updates at all. Users should be patching their systems through installation whenever updates are released.



Behavioral Insights

>>HABITUATION INTENSIFIES UPDATE DEFERRAL. Most systems push update prompts on a relatively regular basis. After a user has deferred updating or acting on a warning multiple times, he or she becomes more likely to habituate that behavior. *Habituation* causes deferral to become a semi-automated process in which the user does very little thinking before hitting the button that corresponds to postponing the update. The correlates to habituation can

also be seen at the level of brain activity: people's neural responses drop after the second exposure to something and continue to decrease with subsequent exposures.⁵

>>THE CHOICE ARCHITECTURE OF MAKING THE DECISION TO UPDATE OR NOT ENCOURAGES USERS TO DEFER. No choice is devoid of context, and these contexts often have an impact on how we decide and act. The modeling of such contexts is sometimes referred to as *choice architecture*—a term coined by Cass Sunstein and Richard Thaler to describe the design of ways in which decision-makers are presented with choices.

The decision of whether or not to update a computer system is no exception to Sunstein and Thaler's framework—the presentation, framing, and contexts of users' choices certainly contributed to the chronic failure to install much-needed security patches. More specifically, updates often require a quick on-the-spot decision: Do I install now or later? If people choose to defer, their sub-decision options are also vague; the user can only choose to be reminded “tomorrow,” “tonight” or “later,” not a more precise time. Such choice architecture unintentionally exploits a handful of heuristic biases to nudge people toward choosing delay every time, making systems less secure.⁶

Additionally, update notifications often come when users are in the middle of something or at other inconvenient times. The prompt to update can interrupt the user's flow and often requires the user to shut down applications she might be using,⁷ which only reinforces a nudge in the wrong direction.

>>PEOPLE MISJUDGE THE LIKELIHOOD OF LOSING PREFERRED FEATURES OR EXPOSING NEW VULNERABILITIES. Stories of patches eliminating features or creating new security flaws are overrepresented in the media and within social networks.^{8,9} When users judge the likelihood of these occurrences, they often employ the *availability heuristic*, a mental shortcut that overweight immediate examples that come to mind and therefore biases prediction about probabilities. Such sensationalized stories can severely distort the view of actual statistics.¹⁰



Design Concepts

1 CLEARLY AND TRANSPARENTLY COMMUNICATE WHAT THE UPDATE

IS DOING. Telling the user what issue(s) the update addresses, how long it will take to install and which features it will modify can help the user understand why the update is critical and can potentially reduce the fear associated with loss of preferred features.^{11,12}

Firms would also be wise to increase transparency in the risks they are helping to avoid. Psychological research suggests that signaling an exertion of effort can increase user satisfaction. In practice, this might entail providing a small pop-up, message, or occasional report that shows users how an update has mitigated vulnerabilities that would have otherwise been exploited (N.B.: Avast software already does this¹³). Such a strategy also makes the importance of updating more salient.

2 REQUIRE AUTOMATIC UPDATES WHENEVER POSSIBLE.

Taking the choice out of the user's hands is one catch-all solution to address the behavioral problems associated with deferral.^{14,15,16} Microsoft is one company infamous for adopting such a strategy.¹⁷ Apple also offers an automatic updating service for apps,¹⁸ and most browsers already automatically update themselves. One thing to keep in mind, however, is that automatic updates should require consent and be installed without impeding the use of the product.

3 PROVIDE UPDATES AT MORE CONVENIENT TIMES.

In situations where requiring updates is not possible, system providers could help combat the negative choice architecture by providing updates at more convenient times. Apple, for instance, gives users the option to update overnight, (between the hours of 2:00 AM and 5:00 AM) when users likely don't need their phones.¹⁹ An extension of this idea would be to have users set a specific time for the update to install in the future, which would allow the user to ensure the update was happening at a convenient time, and pre-commit the user to the update.

SECURITY WARNINGS



The Problem

When browsing the web, users often encounter, and quickly ignore, warnings indicating risks of malware, expired SSL certificates, and untrusted web pages. In fact, users ignore warnings up to 90% of the time.²⁰ By disregarding warnings, we become susceptible to malware that can corrupt or exfiltrate data.



Simple statement of the behavior we want to change

When users receive browser-based warnings telling them that the website they are going to is insecure or has malware, they will often click-through and ignore those warnings, putting themselves at risk. We want users to take security warnings seriously and act appropriately by avoiding websites that have been marked as unsafe.



Behavioral Insights

>>USERS HABITUATE TO WARNINGS. Our neural response to stimuli, including security warnings, drops dramatically after the second exposure and continues to decrease with subsequent exposures.²¹ This phenomenon, called *habituation*, helps to explain why users increasingly ignore similar looking security warnings over time.

>>USERS TREAT ALL WARNINGS AS THE SAME. Clicking through a known malware warning is much riskier than clicking through an expired SSL warning. However, because of habituation, users may treat malware warnings that signal high risk the same as they would benign SSL warnings. The reason users might habituate across different warning types is that the two kinds of warnings often look very similar, and prompt the user with similar actions. Therefore, if a user becomes accustomed to clicking through SSL warnings, they are likely to do the same with malware warnings, despite that the two warning types indicate different severities of risk.²²

>>THE “WARNING WHO CRIED WOLF.” Based on users’ experience with warnings, they may come to distrust the credibility of a warning and choose to ignore it. Many users have received warnings when attempting to access web pages they know to be safe (such as Peter’s experience with the State Department website), and this may prime them to distrust warnings in the future. Additionally, even if malware is present on a web page and gets downloaded onto a user’s computer, the user may never become aware because the malware may act in the background and not produce any meaningful effect on the user’s everyday experience. Without feedback about the real consequences of the malware, the user may choose to disregard warnings in the future because nothing “happened” the last time they did not heed the warning’s advice.

>>AFFECT BIAS AFFECTS YOUR VIEW OF WARNINGS. Users are much more interested in what lies on the web page beyond a warning than on the warning itself. This is particularly the case if the user is expecting something on the other side that they feel positive about, such as a movie they were dying to see, free music from a favorite band, or a baseball game they couldn’t miss. Behavioral science has shown how people’s positive feelings about an experience or an outcome can cause them to discount the risks and inflate the perceived rewards associated with a given action or decision. This *affect bias* can have the consequence of causing people to dismiss warnings when they really should be paying more attention.

>>PRESENT BIAS. The speed at which users access information on the internet puts users into a context in which the present is of much greater consequence than the future. This *present bias* can cause users to overvalue immediate benefits and discount future costs. For instance, users may be overly focused on the immediate benefit of accessing a website, and less focused on the potential and often unclear future costs they would incur from getting malware on their computer today. This context makes it much more likely that users will ignore warnings and continue to the website they were currently interested in seeing.



Design Concepts

1 MAKE THE EXPERIENCE DIFFERENT EACH TIME. One way we might hold at bay the effects of habituation is to develop warnings that vary so that the user experiences something different each time. This could simply mean ran-

domizing a set of distinct user experiences, but it could also mean developing warnings that are dynamic or animated to engage the user. One strand of research in this area has produced the idea of developing polymorphic warnings that change shape, size, and orientation on the screen. Polymorphic warnings have been shown to be far more resistant to habituation, at least in the short term.

2 MAKE THE CONSEQUENCES VIVID. Instead of simply telling the user that there may be malware and informing them that they could have information like passwords stolen, warnings could provide even more vivid information about the potential consequences of losing passwords or other account information. For instance, the warning could alert the user that if their passwords are taken bad actors could break into their email, or if their credit card information is stolen a bad actor could ruin their credit history. Alternatively, warnings could include short videos (30-60 seconds) with real people telling their stories about how they were hacked and what consequences followed (e.g. identity theft, company secrets being stolen, losing their job, etc.). Once the video has played the user would be asked if they still want to proceed. Making salient and vivid the consequences of clicking through a warning may reduce the likelihood that users will dismiss warnings too quickly.

3 MAKE WARNINGS MEANINGFULLY INFORMATIVE. Warnings should not only be more salient, but they should also be clear and actionable to users who are not technology experts. Warning messages should indicate what actions the user should take to return to safety, as well as any additional steps the user can use to protect themselves in the future against similar risks.

4 PROMPT ALTERNATIVE ACTION. One version of this would be to generate “smart” warnings which simultaneously advise the user to avoid the target web page, but also provide alternative web pages with similar content that are not known to have security risks.

5 INCREASE THE HASSLE. Seemingly small hassles, such as extra steps in a process, have been shown to have a disproportionately large effect on behavior. Warnings could leverage this insight by adding additional steps to pass through them such as requiring the user to confirm twice that they want to go through, or requiring the user to wait for 30 seconds or a minute after clicking through a warning before they can proceed.

SAFE CODING



The Problem

According to the U.S. Department of Homeland Security (DHS), 90% of all security incidents result from exploits against defects in software. The attacks on Target²³ and JPMorgan Chase²⁴ as well as the exploitation of bugs such as the now infamous Heartbleed²⁵ and Shellshock²⁶ all occurred because of software defects. Many of these vulnerabilities are attributable to a small set of common programming errors due to inadequate or unsafe coding practices.²⁷ Bugs and code errors are estimated to cost the U.S. economy billions of dollars per year.²⁸

Poor adherence to safe coding practices is widespread. According to Veracode's State of Software Security Report (2015), three out of four applications produced by software vendors fail to meet the Open Web Application Security Project (OWASP) top 10 standards for security. Even more alarming, 63% of internally developed applications are out of compliance with the OWASP top 10 as well.²⁹ While many security professionals indicate that application vulnerabilities are a top concern for them, only a fraction of those professionals says that their companies always scan for potential vulnerabilities during the code development process.³⁰ While important, part of the reason organizations do not prioritize ensuring application security is because doing so is often at tension with getting functional code out into the market.³¹

However, fixing code after release can be exponentially costlier. The Systems Sciences Institute at IBM found that in comparison fixing a bug in the design phase of an application, fixing the same bug in the implementation phase would cost 6.5 times more on average while fixing the bug during the testing or maintenance phase would cost 15 and 100 times more on average respectively.³² Yet, a significant amount of software is released to the public with security vulnerabilities that would later need to be fixed via additional software in the form of a patch. Additionally, some legacy software exists with vulnerabilities that may never get a patch because the organization that built the application has moved on to other projects.

Given the importance of producing safe and secure code and the costs associated with waiting to remediate code errors later in the development lifecycle, why isn't there greater adherence to safe coding practices, including the use of security tools during initial program design?



Simple statement of the behavior we want to change

Many software engineers do not practice secure coding best practices and develop code that contains common programming errors. We want software engineers to practice secure coding best practices and develop code that doesn't contain common programming errors.



Behavioral Insights

>>ENGINEERS TUNNEL ON THEIR IMMEDIATE DELIVERABLES AT THE EXPENSE OF SECURITY. Efficiency is a top priority in software development as exemplified by the shift from 'waterfall' production processes to agile and scrum methodologies over the past two decades.³³ This shift has helped organizations significantly reduce the amount of time it takes to get an application to market, but it potentially comes at the expense of security.³⁴ While there is some debate about whether developing secure code is at odds with these efficient production methodologies, what is clear is that without dedicated attention to safe coding, security can get neglected in favor of producing functional code.³⁵ Part of the challenge is that the speed with which code is developed and pushed out creates a context of scarcity, specifically time scarcity.

Behavioral science has shown that scarcity, or lacking an essential resource (e.g. time, food, money), can place an undue psychological burden on people.³⁶ Imagine human cognitive capacity as a fiber-optic cable. While information can pass through the cable, it has a finite amount of bandwidth, meaning it can only handle a certain degree of information at one time. When a significant portion of bandwidth is occupied, for instance by a large file download, seemingly unrelated activities such as loading a web page or receiving an email take more time.

The human brain functions quite similarly. People only have a certain amount of bandwidth that we spread across all of the various tasks that we need to attend to, and when people operate under conditions of scarcity, it is as if they

are using their fiber-optic cable to download a host of large files. While any one file may not be so large, together they can deplete one's bandwidth to the point that even small decisions and actions can become incredibly difficult.³⁷

The reason for this difficulty is that people tend to "tunnel" in response to scarcity, focusing on tasks and activities that are most urgent. This, in turn, crowds out other actions and considerations that would otherwise compete for attention. When facing the time scarcity imposed by short development sprints, developing functional code is the primary focus of engineers. By focusing on the functional aspects of the code, other details like small security contingencies in the software or errors that do not hinder functionality but open the application up to vulnerabilities may not get the focus that they deserve.

Additionally, time scarcity can exacerbate other undesirable behaviors. For instance, time scarcity may cause engineers to be present-biased, which causes them to avoid immediate costs, potentially at the expense of future benefits. In the context of building a piece of software, checking the code during the programming stage could help an organization avoid the cost of fixing it in the future when it might be more expensive to do so. However, the act of checking the code could feel like a painful hassle that the developer would prefer to avoid to make sure the software is at least functional in time for its release. Even if the developer had an intention to check the code, being present-biased may cause them to decide to put it off until the end of the build instead of stopping throughout to check and fix errors. While in the best case scenario the developer might get around to the task, in the worst they would have overestimated the amount of time they have left at the end of the sprint, and never get around to checking.

>>ENGINEERS USE HEURISTICS WHEN DEVELOPING SOFTWARE THAT MAKES THEM INATTENTIVE TO SECURITY CONCERNs. Building a piece of software is a complicated task. Engineers need to integrate vast amounts of information to make decisions about the code they are writing and frequently rely on their working or short-term memory to keep track of where they are in the code, and how what they have written before might affect what they are about to write next. However, our working memory is limited, which can take a toll on cognitive processes when too much information, options, or decisions are available to them. To get around this, people tend to employ *heuristics*, or cog-

nitive shortcuts used to make decisions and perform tasks that do not require all available information. These heuristics often work well in helping us to navigate decisions and actions in our lives, but sometimes they can cause us to err.

How people go about catching a thrown ball is a good example of how we employ these sorts of heuristics. One way we might explain how people are so adept at catching balls is that our brains are well equipped to solve differential equations on the fly allowing us to figure out where we need to stand and how fast we need to get there to catch the ball. However, what we find in practice is that people instead use a heuristic called the ‘gaze heuristic’ to catch balls. With the gaze heuristic, we may not know where the ball will land, but if we follow the rule that we should gaze at the ball at a constant angle, running faster or slower towards the ball to maintain that angle, we will catch the ball almost every time.

Engineers use similar types of heuristics when developing code. However, often these heuristics are employed with the goal of building software that meets both functional and performance requirements, which tend to neglect security considerations. The reason for this is that security vulnerabilities represent uncommon cases that do not fit comfortably into these heuristics for function and performance. For instance, when engineers build code they need to take into account the possible inputs that it can receive and the states the program can reach. Ensuring that functional and performance requirements are met only requires attending to a set of predictable inputs, and not necessarily those unusual cases that can lead to vulnerabilities and may not be salient to engineers as they build. Additionally, engineers may not consider the security implications of utilizing code that exists within official libraries, third party APIs, and even open source directories like GitHub and other crowd source services Stack Overflow when trying to solve a particular functional problem. That API may present security vulnerabilities, but engineers may use a heuristic that says, “APIs provided by third parties must be safe to use,” and not scrutinize the code to assure themselves that it is safe.



Design Concepts

1 CREATE MORE BANDWIDTH. Engineers are likely already too bandwidth taxed to focus on coding safely while they are simultaneously building functional code. One way to get around this is to find ways to offload the attention

they would need to use to find errors to someone or something else. This could be accomplished in a number of ways. For instance, development teams could include an integrated team member whose sole responsibility would be to vet code for accuracy and safety, and visually identify errors as the development team builds. Another more technical solution would be to create and use better error identification systems. For instance, expanding the error identification capabilities of compilers so that engineers can receive feedback about errors as they go and remediate them in real time. Additionally, organizations could require development teams to utilize existing tools to check for vulnerabilities such as fuzz testing, buffer overflow checks, and network port scanning to name a few. Another way to accomplish this is to build in dedicated time after sprints to vet code before moving on to the next step instead of expecting engineers to budget that time in for themselves.

2 PROVIDE TOOLS TO AUGMENT HEURISTICS. The heuristics that engineers use to solve problems and develop software can help them build functional software faster, but do not require them to be attentive to specific security concerns. One way that we might be able to solve for this is to provide reminders and checklists about specific reflective questions that engineers should be asking themselves that include things like: testing unexpected inputs, thinking through alternative information flows and questioning the security of code lifted from other sources. Providing an opportunity to frame problem-solving decisions in a way that includes security considerations can help improve the heuristics that engineers already use.

3 BRING THE COSTS INTO THE PRESENT. Because the future costs of errors may be diffuse across the organization, engineers may not experience any personal costs associated with those errors. One way to change this would be to bring the costs into the present by creating a performance-based pay model. One way to do this could be to provide bonuses to engineers that make no mistakes, but a more behavioral way to do this would instead be to create costs for engineers when they make mistakes. Behavioral science has shown that people are loss averse, meaning they would rather avoid a loss than accept an equal sized gain. Therefore, it could be possible to promise a set amount of money to engineers after a sprint that reduces for every error they make.

PASSWORDS



The Problem

One of the more discussed behaviors in cybersecurity is users' predictable choice of passwords. Astonishingly, 10,000 of the most common passwords can access 98% of all accounts.³⁸ This recycling generates vulnerabilities because hackers often use dictionaries comprised of common passwords and other frequently used patterns when attempting to crack an account. Even among users who generate strong passwords, however, storage poses additional problems—as many as 50% of people write passwords down in one form or another,³⁹ leaving them open to the obvious attack of someone obtaining a written copy. Some opt for more secure storage methods such as password managers (tools used to store several passwords in one place, gated by a single authentication mechanism). However, usage remains a problem: one survey estimates that only 8% of people use such a manager.⁴⁰ Furthermore, despite the strong security of some password managers,⁴¹ they have been breached in the past.⁴²

Additionally, once a password has been chosen, a user is unlikely to change it unless they receive direct feedback of a breach and if they do make a change, they generally create related passwords that are relatively insecure.⁴³ This habit magnifies the impact of numerous large hacks in which millions of passwords are compromised at the same time.^{44,45,46} The consequences of weak passwords are drastic and wide ranging—over 60% of all cases of identity theft in the U.S. start with compromised passwords. The consequences of poor password construction can extend all the way to leaked classified intelligence or national security documents.⁴⁷



Simple statement of the behavior we want to change

Users do not follow best practices in relation to passwords. Users choose weak and predictable passwords, do not change their passwords as often as they should, and when they do change their passwords, they often reuse common element(s). We want users to follow best practices by changing their passwords often and choosing strong passwords whenever generating new ones.



Behavioral Insights

>>STATUS QUO BIAS INHIBITS PASSWORD AND CHANGES. Before considering how hassles associated with changing a password may affect the decision to switch to a new one, it is helpful to consider the setup and framing of the choice. One way to think about this is that when users receive a prompt to modify a password, they are in essence choosing between sticking with the status quo (i.e. their current password) and deviating from the status quo (i.e. adopting a new password).

Because of this, users become susceptible to *status quo bias*, or an emotional overweighting of the current state of affairs.⁴⁸ In terms of consequential behaviors, status quo bias can cause individuals to disproportionately stick with the status quo.⁴⁹ Here, those behaviors manifest in an irrational tendency to not switch passwords even when prompted to, thereby generating security vulnerabilities.

>>USERS' SYSTEMATICALLY PREDICTABLE THOUGHT PROCESSES HINDER THEIR ABILITY TO GENERATE STRONG PASSWORDS. Faced with requirements to switch passwords or generate ones with special characters or capital letters, users will often comply in *predictable* (and therefore, less secure) ways. One explanation for this behavior is that users generate such passwords in two stages, rather than one. In other words, users will first think of some word or phrase that doesn't include special characters or capital letters and will then incorporate the additional elements retroactively and systematically—for example, putting the capital letter first, placing the special character at the end, or replacing “S” with “\$”.

Additionally, feedback mechanisms such as the green checkmarks that appear as you add additional elements to a new password—symbol, number, capital letter—reward users for compliance, not actual password strength.⁵⁰ Users offload the cognitive work of thinking creatively about a unique password, and instead simply comply with adding symbols where it is easiest. Increasing the types of characters used in a password should increase its entropy, a common measure of password strength, but current prompts fail to nudge users into realizing these benefits. While this approach is better than not including those additional elements, the effect is muted because of the predictability of these

decisions and allows hackers to make a more informed decision about which passwords to test.

Finally, the poor timing of new password prompts—the user is usually focused on using the service or site on which they are attempting to log in—leads to further inattention to the task and only intensifies the problem.

Less secure passwords are easier to remember. Given the practical limits of human memory,⁵¹ it is naturally very difficult for people to remember several different passwords simultaneously. When choosing a password, users often face a choice between one that is easy to remember yet insecure and one that is secure yet hard to remember. In the former case, users commonly reuse passwords or generate insecure variants (see above). In the latter case, users often physically write down passwords, leaving them open to further vulnerabilities.⁵²



Design Concepts

1 INCREASE ADOPTION OF PASSWORD MANAGERS/GENERATORS. A password manager is a tool used to store several passwords in one place, gated by a single authentication mechanism. Often, they also include the ability to generate secure passwords for the user.⁵³

Especially when users have to work with a large number of different systems, using a password manager to generate and store passwords helps alleviate the tension between security and memory burdens. There are some concerns about the security of web-based managers,⁵⁴ especially those that autofill information,^{55,56} and there is always the risk of the password manager itself being hacked (LastPass, for example, was breached in 2015⁵⁷).

That said, opting for a standalone manager that doesn't autofill (e.g. Password Safe⁵⁸) will usually provide more security and convenience for users than attempting to remember several different passwords.⁵⁹ For this reason, despite the risks, using a password manager is generally a good choice for the average user.

2 PROVIDE BEHAVIORALLY-INFORMED HEURISTICS DURING PASSWORD CONSTRUCTION. To reduce the tension between security and convenience, service providers would be wise to provide actionable rules of

thumb into password generation screens and interfaces about how to generate a secure password in such a way that a lay person could grasp. For instance, research suggests that simply increasing the length of passwords is one of the most effective ways to increase password entropy.⁶⁰ Encouraging the use of a ‘passphrase’ (i.e. a semi-random string of words such as “falsefrogpaperbell” or “obesedragonmagentatissue”) would reduce the tug and pull between security and ease of memory.^{61,62} However, it’s important to note that a passphrase does not provide nearly as much security as a string of random letters with the same amount of characters—a good hacker will usually have a dictionary full of common words and phrases at their disposal.⁶³

Perhaps a more secure method would be to tell users to take a sentence and turn it into a password, as Bruce Schneier (among others) has suggested: “This little piggy went to market” might become “tlpWENT2m”. That nine-character password won’t be in any dictionary.⁶⁴ Like nearly any non-random method, this approach also systematically generates some strings of characters more than other and is therefore still less secure than a random string of words.

Despite these caveats, facilitating memorable passwords like “obesedragonmagentatissue” or “tlpWENT2m” through thoughtful UX design would be a marked security improvement for the average user.

3 REQUIRE MORE FREQUENT PASSWORD CHANGES AND INCREASE RESTRICTIONS ON PASSWORD GENERATION. One way to circumvent these behavioral problems would be to leave the user with no choice but to improve his or her password strength. Requiring password changes more frequently or forcing users to use a minimum amount of characters or a baseline of specific characters would improve password security across the board. This could potentially be enforced through regulation or through service providers themselves.

If one were to adopt such an approach, however, a number of concerns would have to be considered. For one, as outlined above, requiring new characters doesn’t guarantee the security one would hope given the predictable ways in which they would be added. It could also potentially annoy users, which would make the approach a harder sell for service providers.

That said, getting users into the habit of changing their passwords could help assuage the annoyance concern over time—people’s neural responses generally drop after a second exposure to something and continue to decrease with subsequent exposures.⁶⁵ Habitually changing passwords might also combat the negatively reinforcing loop of status quo bias (see above) by reducing the amount of time users have any one password.

4 ELIMINATE PASSWORDS ALTOGETHER. Given the existence of more secure, alternate authentication mechanisms—such as biometric authentication—one catchall solution to the behavioral problems associated with passwords would be to eliminate them entirely and to replace them with mechanisms less susceptible to human error.

Many have expressed support for such a strategy. For example, Netscape co-founder Jim Clark has been a vocal proponent of this,⁶⁶ and in one survey, 84% percent of respondents (taken from the general population) were in support of completely doing away with passwords.⁶⁷

If passwords were eliminated, however, they would need to be replaced with a better alternative. Multi-factor authentication offers one solution and some firms are exploring whether bold ideas such as digital tattoos or password pills could be used in the future as identification and authentication mechanisms.⁶⁸ In the short term, Google’s trust API (codename: Project Abacus) could be one viable replacement—it uses a mix of multiple weaker indicators that, together, can create a stronger, more convenient authentication mechanism.⁶⁹

MULTIFACTOR AUTHENTICATION USE



The Problem

Given concerns about the efficacy and security of passwords, the availability of multi-factor authentication (MFA) methods has been increasing. However, despite the added security that MFA can provide to consumers and organizations, there are still low rates of adoption. Among Gmail users, adoption of two-factor authentication likely hovers around 10%, though Google has not made the actual data public.⁷⁰ Dropbox reported to Brian Krebs that of their 500 million users, only 1% has adopted two-factor authentication despite its availability over the past four years.⁷¹ While the most common forms of multifactor authentication are far from a panacea, any additional layers of security on top of passwords should always be considered.



Simple statement of the behavior we want to change

Users do not activate and use multifactor authentication when it is provided; we want users to always use multifactor authentication when it is provided.



Behavioral Insights

>>TWO-FACTOR AUTHENTICATION IS A HASSLE TO BOTH SET UP AND USE. Behavioral literature suggests that small hassles and even the perception of small hassles can create severely disproportionate consequences.⁷² In order to use multifactor authentication (MFA), users must go into their security settings and turn on the service. However, the process of doing so may not be clear to the user, causing them to either delay or put off setting up MFA indefinitely. Another hassle that might cause users to either abandon or not use MFA in the first place is that using multifactor often requires some additional piece of hardware to deliver an authentication code. The most prevalent form of MFA is SMS-based two-factor authentication where the authentication code is sent to a registered phone number. Users may forget to charge their phones, or otherwise be out of coverage areas when they need access to their accounts, which could make using multifactor authentication feel like too much of a hassle.

>>OPT-IN STRUCTURE AND STATUS QUO BIAS DETER USE. Unless the workplace administrator requires multifactor authentication to be used, these sorts of authentication tools are often opt-in services for employees. Behavioral science has shown that when things are opt-in by default, people are less likely to use them than if they are opt-out instead. This is because of a phenomenon called *status quo bias*, which describes our human propensity to stick with the status quo as opposed to changing it.⁷³ Therefore, when the status quo doesn't include MFA, it is likely that users will keep it that way.

>>OVERCONFIDENCE IN PASSWORDS INHIBITS ADOPTION. Some users may see multifactor authentication as a "nice to have" but not necessary measure, believing either that their password alone is enough to keep them safe, or that they doubt they would ever become the target of an attack. Part of this could be due to the fact that many websites provide some feedback about the quality of a password as it is initially chosen, which may lull users into a false sense of security once they enter a "secure" password. Other users may be unaware of the likelihood that any password can be broken. This context can make people overconfident in how safe they are without MFA, and therefore less willing to adopt it.



Design Concepts

1 DEFAULT PEOPLE INTO USING MFA. One way to get users to adopt multi-factor authentication is to change the default. Instead of defaulting people into a situation where they need to opt-in to MFA, providers could default new users into activating MFA and providing them an opportunity to opt-out if they don't want to use it. If defaulting people into MFA is not feasible, another alternative could be to force users to make a choice during service onboarding about whether they want to use MFA, this way they can't avoid the decision altogether.

2 PROVIDE MULTIPLE AUTHENTICATION METHODS, AND MAKE REQUIREMENTS OF USE CONTEXTUAL. It might be possible to reduce the hassles involved in signing up for and using MFA by providing a number of different ways that users can adopt it. SMS-based two-factor authentication may not be the best option for all users, and by providing a number of different methods, users can better select those that are most comfortable for them. Additionally, it may be possible to only require users to utilize MFA based on

context. For instance, if a system can determine that a user is logging in from a work location at a time that they normally log in, then the system can remove the authentication requirement. However, if the user is logging in from an unusual location or at an unusual time then MFA can kick in to ensure the user is who they say they are. This way it will be possible to remove the hassle of MFA at times when hassles might be least welcome or necessary.

3 PROVIDE VIVID FEEDBACK ABOUT SECURITY OF INCLUDING MFA.

Current services may provide some indication of how secure a password is, but they do not go so far as to provide a feedback metric about security generally. One could imagine that when users are setting their passwords, the system could provide some visual feedback that the password is strong, but that the user has only achieved 50 percent of the possible security on the account, and give them the option to increase that security to 100 percent by adding MFA. By doing so, it might be possible to reduce the overconfidence that some users might have about the security their password alone brings.

EMPLOYEE COMMITMENT TO CYBERSECURITY



The Problem

As of 2016, according to Symantec, one in every 196 emails has a virus, and one in every eight websites has a critical, unpatched vulnerability.⁷⁴ In an enterprise setting, it's easy to see how one absent-minded click can lead to a multi-million-dollar breach. Despite the critical importance of widespread adherence to data security procedures, as many as half of all U.S. workers aren't sure if their organization even has information security policies.⁷⁵

Traditionally, technical controls have been used to enhance information security, with less emphasis given to the human side of security.⁷⁶ However, academics and practitioners alike have begun to realize that security cannot be achieved without the support and engagement of the people involved.



Simple statement of the behavior we want to change

In addition to behaviors that are intentionally malicious, like breaking into a company's secure files, users also make naïve mistakes, like choosing weak passwords and using insecure hardware like flash-drives.⁷⁷ Even in organizations that mandate information security training, users may fail to exhibit the type of awareness and vigilance that their trainers expect. We want end users to uphold security policies to protect the integrity of company data and systems.⁷⁸



Behavioral Insights

>>**EMPLOYEES DO NOT FEEL ENGAGED AND COMMITTED.** Employee engagement is central to a company's ability to protect and secure information. Organizational commitment describes an individual's sense of attachment to their organization. Factors such as work stress, bandwidth, work-life balance, mastery, and autonomy can all impact a person's level of psychological attachment to their organization.⁷⁹ Committed employees tend to engage in activities that are beneficial to their organization because they believe that their actions

will improve organizational outcomes. Employees embedded in workplace cultures that foster trust and mutual ownership are more likely to have an interest in and intention to comply with cybersecurity policies, as well as to display behavior that goes above and beyond the call of duty.

>>THE BEHAVIOR OF OUR PEERS AFFECTS OUR BEHAVIOR. It's human nature to follow others. For better or for worse, we are influenced by what we perceive other people are doing, even when that perception is wrong. This phenomenon is known as *social proof*.⁸⁰ In any given situation, we assume that the people around us know what the correct behavior is, and therefore, we adjust our behaviors accordingly.

In the context of cybersecurity, observing people's security behavior can be quite difficult. For instance, rarely does one ever know whether their colleagues or friends are actively using two-factor authentication, or applying secure password generation methodologies.⁸¹ However, if and when the security behaviors of others are visible, the visibility of those behaviors can have a significant impact on users' adoption of various security measures.⁸²

>>PEOPLE THINK THEIR ACTIONS DON'T MATTER. Individuals are less likely to take responsibility when others are present. Psychologists call this phenomenon when people reduce their effort when working in groups, *social loafing*.⁸³

This diffusion of responsibility is partially driven by the fact that an individual assumes that if taking action is important, someone else would have done so already.⁸⁴ When cyber responsibilities are not explicitly assigned, they can be neglected altogether.

Further, a person's decision to act can be in part rooted in their perceived ability to affect an outcome or make a difference. If end users believe that their actions make little difference in achieving overall security, they may be less likely to follow security policies.⁸⁵

>>SECURITY ISN'T TOP-OF-MIND. Decades of behavioral science research have demonstrated that people have limited attention and face competing demands for their mental bandwidth.⁸⁶ This limited bandwidth leads us to selectively concentrate on one aspect of the environment while ignoring other

aspects. Because of our limited bandwidth, we might not attend to security behaviors that matter, pay attention to guidance that is salient but possibly less relevant, or not attend to security responsibilities on time.

Employees have many pressing concerns, including their jobs, families, and social lives. Often, the security of the office computer just doesn't make the cut. Research has shown that focusing on an unmet need, such as money or time, can impede our ability to focus on other things. Living in a condition of scarcity taxes our cognitive capacity and executive control, diminishing intelligence, impulse control, problem-solving ability, and more.⁸⁷



Design Concepts

1 Ensure employees are engaged. Satisfied employees generally feel a ‘sense of oneness’ with their organization.⁸⁸ These same employees may psychologically perceive a threat to the organization as a threat to the self. Such employees may be more likely to engage in safe end user security. For employees that regard computer security as an obstacle to productivity, awareness should emphasize how security, from a broader perspective, contributes to productivity.

2 Show what others are doing. In many cases, simply showing someone what others are doing can change their behavior. One way to do this is to make compliance visible by telling employees what their peers are doing. For example, management could send a monthly email highlighting the number of people that have updated their security settings. It's important to note, however, that such cues are only useful if a large subset of a group is engaging in the “good” behavior.

3 Make security top-of-mind. Studies show that simple reminders can be incredibly valuable in helping people follow through on their intentions.⁸⁹ Reminders are simple yet powerful tools that can draw people’s attention to the right task at the right time. Reminders work because they can make an otherwise non-salient task, topic, or item top of mind.

ACCESS CONTROL MANAGEMENT



The Problem

In an enterprise setting, ensuring the right people—and only the right people—have access to appropriate data and resources is a critical piece of information security. Poorly designed controls and outdated access control settings inadvertently create attack vectors, allowing potential hackers to use a former employee's still valid account to access sensitive information.

In general, the term *access controls* refers to the ways user accounts are managed, the manner in which privileges and entitlements to information or computing resources are granted and rescinded, and the overall design of the permission architecture, be it a tiered waterfall design or a compartmentalized approach.⁹⁰ We suggest that IT administrators and security professionals take a broader view of access controls and consider all the possible ways their system can be accessed, including the physical security that should prevent an unauthorized person from entering a workspace.

The failure to actively and effectively manage access controls has likely been a major factor in some of the most well-publicized data breaches. For example, the hackers who stole the personal data of about 80 million Anthem customers had at least five sets of valid login credentials, likely obtained through a phishing attack. Using these valid account credentials, the hackers then queried Anthem's database in ways that should have been "unauthorized," but the account entitlements had been improperly set or updated and thus the hackers were able to execute queries that according to Anthem's policies were authorized.⁹¹ In a separate, well-known case at Target, poorly designed controls (no white-listing of software on the point of sale machines and the connection between an internet-facing vendor system, which was exploited using an HVAC vendor account to other business critical systems) failed to stymie hacker efforts.⁹²



Simple statement of the behavior we want to change

IT and security administrators fail to improve the design of an access controls scheme and fail to appropriately maintain and update entitlements in whichever system their organization is using. We want IT and security administrators to properly maintain and update entitlements on a regular basis and in compliance with their organization's policies.



Behavioral Insights

>>IT ADMINISTRATORS AND INFORMATION SECURITY SPECIALISTS

JUGGLE MANY RESPONSIBILITIES. Burnout has become a prominent concern in the information security field.⁹³ Recognized as a cost center for enterprises, security teams and IT administrators can wind up under-resourced for what is becoming increasingly demanding work. Stories abound of IT personnel having too much on their plates. In this type of environment, individuals may wind up with scarce time resources to address issues that are perceived to be vitally urgent and have little time or energy left to execute mundane tasks associated with managing access controls such as removing old users and updating whitelists.

Behavioral science has shown that the context of scarcity, or lacking an essential resource (e.g. time, food, money), can place an undue psychological burden on people. Imagine a person's cognitive capacity as a fiber-optic cable. While information can pass through the cable, the cable has a finite amount of "bandwidth," meaning it can only handle a certain volume of information at one time. When a significant portion of bandwidth is occupied, for instance by a large file download, seemingly unrelated activities such as loading a web page or receiving an email take more time.

The human brain functions quite similarly. People only have a certain amount of bandwidth that they spread across all of the various tasks that they need to attend to, and when people are operating under conditions of scarcity, it is as if they are using their fiber-optic cable to download a host of large files. While any one file may not be so taxing, in concert, they can deplete one's bandwidth to the point that even small decisions and actions can become incredibly difficult.

The science of scarcity suggests that in situations where individuals have lim-

ited cognitive bandwidth, they are prone to experience several psychological phenomena that impair decision-making and follow through.⁹⁴ First, they will tend to “tunnel” in response to scarcity, focusing on tasks and activities that are most urgent, which in turn crowds out other actions and considerations that would otherwise compete for attention. For instance, if an enterprise’s IT or security team’s focus is directed on urgent tasks like identifying existing malware, or evaluating a new security-related investment, the day-to-day grind of managing controls may continually get postponed. This “tunnel” effect will also make IT administrations less likely to reach out to end users for guidance what privileges they may need, even though the end users are likely more familiar with resources and more likely to identify their resource needs.

>>IT AND SECURITY ADMINISTRATORS MAY ASSUME THEIR ACCESS

CONTROL SCHEME IS SAFE. The phrase ‘if it ain’t broke, don’t fix it’ seems especially misplaced when describing an enterprise access control system. Just because the scheme hasn’t failed yet doesn’t mean it won’t fail tomorrow. Psychologist Peter Wason identified a phenomenon called the *congruence bias* in which people will mentally repeatedly test a single hypothesis without evaluating alternative hypotheses.⁹⁵ In this case, if an access control scheme appears to be serving its purpose, enterprise decision-makers will accept that the system works and won’t consider the myriad of ways it could become compromised. This effect demonstrates the importance of thinking about enterprise cybersecurity holistically. However, if decision-makers ask narrow questions about the immediate effectiveness of access controls, they are apt to make less secure choices—e.g. accepting an access control scheme ‘as is’—than to be continually searching for ways to improve security.

>>IT ADMINISTRATORS AND INFORMATION SECURITY PROFESSION-

ALS FORGET THE LITERAL BACK DOOR. Even with robust policies and top-notch technical safeguards, an enterprise can still be breached the old-fashioned way: a team member steps outside to make a personal call, leaves the door ajar, and suddenly a USB flash drive with malicious code is inserted into a networked computer. IT and security professionals may not think of themselves as being responsible for building security, but the failure to include physical access control in a well-developed access control scheme leaves an additional vulnerability. This oversight may be due to a *mental model* of security that views cybersecurity as a chiefly technological problem that must require

solely technical solutions. We all use mental models to both categorize and make complex and abstract things easier to comprehend and use. While they are very useful, as the scenario above makes clear, mental models can also be limiting. With Verizon reporting that nearly 40% of all physical thefts of laptops and hard drives occur in the workplace, a limited mental view of access controls can leave a critical vulnerability unaddressed.⁹⁶



Design Concepts

1 Automate the entitlement review process. By automating parts of the process, we can limit the amount of cognitive bandwidth required to maintain and manage access controls. An automated process can also leverage defaults, such as entitlements that automatically expire after a given amount of time.

2 Provide planning prompts and timely reminders. Setting clear moments when entitlements must be reviewed and must be updated—followed by timely reminders for each—can reduce the cognitive bandwidth necessary to take in-the-moment action to update access controls.

3 Empower all team members to set access controls by creating “Break the Glass” options. End users likely know best which files and resources they need and which they don’t need. But asking for permission, getting approval, and then getting someone to change the control setting is a hassle, and if the IT administration isn’t available, the end user is out of luck. It is important to trust team members to make sound judgments and give them the ability to access resources for a period of time without review.

4 Leverage inexpensive changes to the physical surroundings of your device. Expensive physical security (e.g. guards or sensors) systems may not be as effective as simple but strategic tweaks to make the workspace more secure. For example, on a team that regularly reviews classified information, but whose members have no need for a camera on their company-issued phones, simply destroy the camera on the phone with a drill bit. Even if the phone’s software is compromised, an attacker won’t be able to use the camera to view any files or documents the owner may be reviewing. Getting out of the “security needs a technology solution” mindset, and seeing alternative, non-technical ways of solving problems, can be highly effective, and significantly less complicated than provisioning some new technology.

THREAT AND VULNERABILITY SHARING



The Problem

The volume, complexity, and sophistication of cyberattacks have increased substantially over the past few years, partly driven by increasing coordination and sharing among otherwise disparate hackers and hacker groups via the Internet.⁹⁷ Some believe that for public and private organizations to respond promptly to these threats and remediate vulnerabilities before major attacks occur, it is imperative for them to participate in the sharing of threat and vulnerability intelligence with one another and coordinate responses.^{98,99,100} In a recent threat intelligence report, 91% of surveyed cybersecurity professionals expressed interest in using cyber threat intelligence¹⁰¹ highlighting the demand for and value of this kind of information. While the U.S. government, NGOs, and private organizations have been working to establish threat and vulnerability sharing frameworks, institutions (CERT, ISACs, NCCIC, NCJTF, OTX, XFE, etc.), standards (STIX, CybOX, TAXII, CVE, etc.), and the necessary regulatory environment to promote their use (CISA)¹⁰², there remain significant challenges to achieving the levels of participation and cooperation necessary for these institutions to provide the anticipated benefits.

For example, of those participants surveyed in the threat intelligence report, only 24% stated that they were “very likely” to share their own vulnerabilities.¹⁰³ In some sense, lack of participation is a purely economic problem. Private institutions have justifiable concerns that the sharing of threat information both within their industry and with the government will not come without costs. Specifically, many for-profit firms are concerned that if they were to share they might come under public scrutiny for their vulnerabilities, that their competitors would gain an advantage over them, or that sharing could make them liable for any consequences of an attack. All of these scenarios could have an impact on stock prices.¹⁰⁴ While the government has tried to reduce some of these concerns by crafting legislation that would protect industry from liability when sharing, critics argue that these measures are incomplete.¹⁰⁵

However, while there are significant economic disincentives to participation and sharing, there could also be other non-economic disincentives that could also be motivating non-participation.



Simple statement of the behavior we want to change

Private sector organizations are not participating in threat and vulnerability sharing within established national frameworks. We want private sector organizations to participate in threat and vulnerability sharing within established national frameworks.



Behavioral Insights

>>PRESENT COSTS DOMINATE COOPERATION DECISIONS. Sharing in the cybersecurity ecosystem is inherently about cooperation. Specifically, if an organization cares to reap the benefits of others sharing their threat and vulnerability information, that organization should participate and share as well. However, what we find is that some private sector firms may not share despite the fact that everyone would be better off if there were full participation. Part of the reason has to do with the fact that the benefits of sharing are likely felt in the future, while the potential costs of sharing could be borne almost immediately. For example, if sharing results in an organization revealing that they had a data breach, that organization might have to pay penalties immediately because of the data breach. Moreover, there were no immediate benefits to sharing. Instead, the organization would likely need to wait to receive those benefits. Behavioral science can help explain why it is sometimes so hard to be patient. *Present bias* describes our tendency to let immediate costs far outweigh future (and potentially greater) benefits. In the case of sharing, both the potential immediate costs and risks associated with sharing likely outweigh the future benefit, which might cause a firm that is otherwise interested in sharing just to say, “we’ll do it next time.”

>>PRIVATE SECTOR ORGANIZATIONS OVERWEIGHT THE LIKELIHOOD OF COSTS ASSOCIATED WITH SHARING. While there may be costs associated with vulnerability sharing, organizations may overweight the likelihood that those costs will occur, as well as the magnitude of those costs. One reason why that might happen is because of *availability bias*. The availability bias describes how our assessment of probability can be affected by our ability

to recall instances of something occurring in the past. We think things are more likely to happen if we can remember cases of them happening easily. However, just because we can remember something happening, doesn't mean it occurs frequently. For example, people tend to believe that shark attacks are more likely to happen than they actually are. Some have speculated that movies like Jaws and events like Shark Week on the Discovery Channel make shark attacks more salient in people's minds, which causes people to believe that shark attacks are more prevalent than they actually are. In the context of sharing, what might be top of mind for organizations is that when they hear about another cyberattack from the news, they recall that the firm lost share value as a result of the attack. What may not have been reported, and therefore what is salient, is that most of those firms recovered completely within a short period of time.¹⁰⁶ What happens instead is the likelihood of losses is perceived to be high, while the likelihood of a quick recovery is thought to be low. Had firms weighed the likelihood of recovery accurately, they may have been more inclined to share because they would have perceived the short-term risks to be lower.

>>ENTERPRISES DO NOT KNOW HOW OR WHERE TO SHARE THREAT

AND VULNERABILITY DATA. While sophisticated enterprises participate in ISACs, and their information security teams may regularly exchange information through personal networks, many enterprises simply do not have, or do not know, through what channel to share cyber threat and vulnerability information. Additionally, even if they did, there may be other associated hassles that could stop them from participating. We often underestimate how small hassles can get in the way of someone following through on their intentions. In the context of sharing, small hassles like filling out long forms, being unsure of what kind of information to include, or translating collected information into some current standard could cause someone to defer reporting on behalf of an institution, potentially indefinitely.



Design Concepts

1 Make the benefits of information sharing more visible. Highlight what other firms are doing and what has come from their actions. For example, enterprises could share the good news of attacks thwarted or vulnerabilities patched. By making the benefits and consequences of sharing vivid, organizations may cease to overweight the immediate costs of the activity.

2 Reduce the upfront costs associated with sharing. While organizations may be overweighting the costs and risks associated with sharing, so long as they are making a decision between incurring a cost and not incurring a cost, they will always be less likely to participate, regardless of the quantified benefit due to *loss aversion*. By developing a regulatory or legislative environment that better protects firms from liability and anonymizes their participation (to protect from public opinion), the decision to participate will be much easier. Focusing on only increasing the salience of the benefits of participation will likely have less effect.

3 Make reporting easier. Improving reporting could be as simple as reducing the hassles involved. With the current growth of both reporting institutions as well as standards, it is likely confusing for less sophisticated organizations to figure out how they can participate. By developing well-defined reporting standards, standing behind a single reporting infrastructure, and providing clear and actionable information about how organizations can participate as well as when and what they should be reporting, it may be possible to increase the participation of organizations who would like to volunteer, but for whom the small difficulties in the process get in the way.

END USER SECURITY SETTINGS



The Problem

Operating systems, popular web-based services, including social media sites like Facebook, and some IoT hardware offer users the opportunity to set and modify settings that can impact users' privacy and security. The design of these settings interfaces (by software or hardware companies) and their management by end users have profound implications for the security of the user's personal information, and in turn, any enterprise of which the user is a member.

Hackers can use information on social media to take a better guess at passwords or to set a personalized spear-phishing email. Thieves can see vacation dates and know when to rob a home or business. Bad actors can leverage insecurities in Wi-Fi enabled devices to construct intricate botnets or spy on unsuspecting individuals.

To keep users safe, end users need to understand, maintain and periodically update security and privacy settings. While the availability of security and privacy settings may well give each user the freedom to 'control' her personal information or access to hardware, in practice, users may adopt (either consciously or unconsciously) settings that are less secure than they intend or would prefer. For instance, in one study of Facebook users, researchers reported a gap between privacy intentions and actual behaviors (settings and sharing) for every study participant.¹⁰⁷



Simple statement of the behavior we want to change

Users do not change default settings, and rarely review or alter settings that affect the privacy and security of their personal information and the devices they use. We want users to be aware of and attend to their personal security settings across all devices and services they use.



Behavioral Insights

>>USERS TEND TO KEEP CURRENT OR DEFAULT SETTINGS. All people tend to have an emotional preference for the way things are, known as *status quo bias*. Status quo bias is one reason users are not likely to shift away from default settings. This reluctance to change the setting is agnostic to the settings themselves, whether designers and developers design those settings for security or to facilitate openness. This user behavior highlights the importance of *defaults*. If the settings are less-than-secure by default when users first begin using a product, users are not likely to change the privacy or security configuration. How software providers set the defaults has a powerful influence on the overall security of user data and the hardware they use. The idea of designing a default is at the core of one of the most famous applications of behavioral science: defaulting employees into retirement plans and pre-setting contribution escalations.

>>THE DESIGN OF SETTINGS OPTIONS MATTERS. Users need to be able to find, navigate to, and then understand settings options. When users can't find the settings or when users don't clearly understand the settings options, they are less likely to make changes. Security researchers have shown that while users have a general, high-level understanding of the importance of privacy settings, they view adjusting the privacy settings not as an essential step to protect against hacks, but as a type of cost for using the 'free' service. Through this lens, users have a limited amount of 'goodwill' for trying to figure what the appropriate privacy settings should be, and there is a drain on that goodwill when the settings options are difficult to understand.¹⁰⁸ One explanation offered by behavioral science for the evaporation of user goodwill is the idea of *choice overload*.¹⁰⁹ Studies on shoppers in the market for consumer products have shown that even if a customer states a preference for maximum freedom of choice when that customer has myriad options presented to them, she becomes demotivated and is less likely to purchase a product at all.¹¹⁰ In the context of modifying settings, if choosing the right settings is hard, users may avoid taking action.

Additionally, a user's engagement (or lack thereof) with the security settings during their first user experience and the visual layout of those options influ-

ences whether users will be attentive to settings in the first place, and which options users are more likely to select. The manner in which options are presented and arranged is sometimes referred to as *choice architecture*. Limiting the number of settings and options a user must choose from is one way a designer may alter the choice architecture in order to avoid choice overload.

It's safe to assume that popular, well-capitalized platforms, especially social media, have invested heavily in defining both the security settings, the options available within each setting, and studied how changes in choice architecture influence users' settings choices. This type of data and continued research is a significant step in learning how best to design small nudges that help users make secure settings choices.

>>INSECURE DEFAULTS MAKE UNSAFE BEHAVIOR TOO EASY. An additional aspect of settings design is the way insecure defaults can unwittingly promote insecure user behavior. The primary example is an operating system default to join an open Wi-Fi network, a behavior that is known to risk personal data. The default to an open network is what psychologists call a *channel factor* or something that makes it easier for someone to maintain their current intentions. A user wants to get online; the default to connect to an open Wi-Fi network makes following through on that intention easy. As a general rule of thumb, settings should be designed to make secure behavior easier and insecure behavior harder, not the other way around.

>>USERS UNDERESTIMATE THE RISKS OF SHARING PERSONAL INFORMATION. Having a mental model or feeling of ‘*who am I? No one wants my data*’ is reflective of a tendency towards overconfidence. Users are assuming the probability of not being hacked is considerably greater than it actually is. Two key contextual features likely lead to overconfidence in these instances. First, users often underestimate how much information they intentionally share online and are often unaware of how much data is available to be captured through their internet-enabled devices.¹¹¹ Second, users don’t realize how the data they share online over time and on different software platforms can be aggregated by hackers to gather a fairly robust portrait of who the user is offline.



Design Concepts

1 Force choice around secure defaults. Flipping the default from insecure to totally secure could have a significant positive effect on user security, but may not be feasible, or even preferable for the user in all circumstances (e.g. it could hinder the usability of a service or a device to an unnecessary or unfavorable degree). Instead, online service providers, software developers, and device manufacturers should start with stringent default settings, and then force users upon the first interaction with the product or service to set the security settings to their own preferences. By doing so, product and service providers can avoid users' status quo bias, and provide a moment of action for the user to think critically about their security preferences.

2 Standardize privacy measures and settings across services. In the world of food, whether it's cereal or chips, the nutritional information on a package label is formatted identically, making key measures—calories, saturated fat—easy to find. Like foods, software services vary widely, but there are likely a few measures, such as the total number of people who can see a post shared on social media, which could be easily developed and standardized across services.

3 Provide users feedback on overall privacy and security. While software or services may provide users with many settings, each of which is modifiable in different ways, the overall security of a user's account may not be salient to them. A single salient metric could take into account information from privacy settings, password strength, and the user's behavior (e.g. logging on from an open Wi-Fi network) and give users meaningful feedback on the security of their account. Additionally, service providers could also give users actionable pointers about how they can remediate insecurities when they arise, giving users an opportunity to improve their overall security score.

4 Leverage social norms. Social media platforms that store significant amounts of personal information and facilitate social networks for users can take advantage of the important signaling function of social norms by showing a user the relative security of their peers, or information about their most secure peers. For instance, when sending users notifications about updating their security preferences, service providers could include information

about the number of close contacts who have recently updated their security preferences.¹¹² Additionally, once users look at their security preferences, service providers could give users information about the number of close contacts that utilized specific security features. This intervention could also include the concept of providing clear feedback, as described above, by using a standardized metric of comparison across users.

PHISHING



The Problem

In 2016, users experienced the highest number of phishing attacks ever recorded. Over 1.2 million attacks were registered in 2016 by the Anti-Phishing Working Group, a global coalition of law enforcement agencies, representing a 65 percent increase in registered attacks over the previous year.¹¹³ Awareness does not appear to be the deciding factor, as users still click on malicious links and downloads despite knowing the risks.¹¹⁴ In fact, many of the most sophisticated and damaging cyber attacks begin with a well-executed spear-phishing attack. Nearly two-thirds of IT decision makers say that spear-phishing is their top concern,¹¹⁵ and in testimony to Congress, the CEO of Fire Eye stated that, since 2015, 95% of the breaches they've remediated began with a spear phishing email.¹¹⁶



Simple statement of the behavior we want to change

Users click on malicious links in emails that spoof or mimic banks, technology companies, coworkers, or any social/professional affiliation of the user. The link itself may initiate installation of malware, may lead the user to a fake (but familiar looking web page) to capture the user's credentials, or the user may unwittingly reveal information by corresponding with the sender. We want users to avoid clicking on malicious links sent via phishing attacks.



Behavioral Insights

>>PEOPLE COMPLY WITH REQUESTS FROM AUTHORITY FIGURES.

When individuals with authority make requests, be it in person, via email, over the phone, or through any other medium, people have a tendency to comply.¹¹⁷ Bad actors perpetrating phishing attacks use this insight to get their unwitting victims to disclose information or download malware onto their computer by masquerading as a person of authority, such as a supervisor, professor, doctor, or another figure with perceived influence. By just using an authoritative title, phishing attacks can trigger a quick-acting *heuristic*, or mental shortcut for deciding how to act, which causes people to equate a request from a person of authority as something with which they should comply.¹¹⁸

>>PHISHING PLAYS ON FAMILIARITY. Familiar people, experiences, and institutions can engender feelings of trust in individuals.¹¹⁹ However, in the virtual world, it is very easy for bad actors to copy the visual and experiential cues that individuals find familiar, such as corporate logos, web pages, and the names of friends and family. By presenting familiar cues to the user, bad actors can build a façade of legitimacy, and lead users to do things they shouldn't do such as download malware or disclose personal information.

>>PHISHING EMAILS PRESSURE USERS TO ACT QUICKLY. Phishing emails are often crafted to create a sense of urgency for the targeted user. By using trigger words such as “alert,” “urgent,” or requesting that the user responds or completes a task “ASAP,” attackers can prompt users to think and act too quickly, making it less likely that they’ll notice that they’re falling into a trap.¹²⁰ Part of the reason creating a sense of urgency might be effective is because people are loss averse, and will do what they can to avoid losses where possible.¹²¹ If people perceive that they’ll lose something if they don’t act quickly, they may be more prone to act without thinking.

>>PHISHING EMAILS AND SPOOFS TAKE ADVANTAGE OF OUR LIMITED ATTENTION. Phishing emails and spoofed web pages almost always contain information that can indicate to the user that the email, attachment or web page is malicious (e.g. pixelated images, slightly different URLs, etc.). However, users may not always be attentive to those details because of *limited attention*.¹²² Attention, much like a limited resource, gets depleted when in use. For instance, if a user directs their attention to some aspect of a user interface, they will have less attention to direct to other details. Additionally, bad actors executing phishing attacks can *prime* their victims to be attentive to specific details, while simultaneously directing their attention away from cues that would signal that the email, website or attachment may be malicious. For instance, bad actors might send an email asking someone to log into their account, priming the victim to be more focused on the login interface than other cues of malicious intent such as URLs, or pixelated graphics.

>>PHISHING EMAILS EXPLOIT OUR CURIOSITY. Pictures from a party, a sample contract from a competitor, celebrity gossip—sometimes the desire to look obscures a user’s ability to weigh the likelihood that the email may be a phishing attack. In at least one study, researchers triggered a person’s curi-

osity by using traditional ‘reverse psychology,’ suggesting that the recipients received a message in error and should not click on a link to an external photo-hosting website. In another example, researchers drafted phishing emails that appeared to present recipients with a personalized opportunity, such as a journalist wishing to write about the recipient’s work with a convenient link to the reporter’s previous writing.¹²³ By exploiting peoples’ desire to close the *curiosity gap*, bad actors can manipulate users into clicking on links and downloading files that they shouldn’t.



Design Concepts

1 Provide real-time feedback. ‘Just-in-time teaching’ can help users connect actions to consequences, eventually pausing the ‘fast thinking’ that characterizes so much email behavior. Researchers have already shown how real-time feedback and just-in-time training can be effective at teaching users how to identify and avoid phishing attacks and website spoofs in real-world environments.^{124,125} Organizations interested in reducing phishing rates should consider adopting these sorts of tools across their enterprises.

2 Slow user reactions. To make users more attentive to the little cues and details that characterize phishing attacks, UX designers could build interfaces to help users ‘slow’ their thinking. While slowing down users may be in conflict with the productivity goals of organizations, but it may be a necessary step in improving enterprise security. One way to accomplish this might be to embed small hassles into the email user experience. For instance, when clicking on a link or file within or attached to an email, the user could be prompted, via a pop-up, to consider whether the link or attachment is from a trusted source. If the user is unsure, an available call to action could be used to quickly and easily send a confirmatory email back to the sender. Slowing down the user in such a way could improve their identification of malicious emails.

3 Reward savvy behavior. Recognize employees who pass sophisticated phishing tests or catch an actual spear-phish with public recognition or financial incentives. While pure incentives are not inherently behavioral, well-constructed incentive programs can have the added effect of getting users to be more attentive to the details of emails, making it more likely users would catch potential phishing attacks before they occur.

4 **Adjust cultural norms through rules of thumb.** Develop organizational policies that disallow sharing of links or attachments through email to avoid any ambiguity when a potentially malicious link or attachment shows up. Instead, provide employees with new platforms and rules of thumb about how to send links and attachments to colleagues through other enterprise services. Additionally, if and when a link or attachment appears in an email in-box sent from a fellow employee, establish heuristics that guide employees to ask the sender about whether they had sent the link or email intentionally. Simply asking, “Hey, did you send me this?” can be the difference between a successful attack and one avoided.

5 **Add more information to URL address bars.** By mixing up colors and mixing in words and padlocks, a web browser can purposely recapture a user’s attention and focus, thus increasing the likelihood that the user will spot a spoofed URL.

INVESTING IN CYBERSECURITY



The Problem

While cost estimates vary widely, there is widespread agreement that data breaches and hacks represent one of the largest risks to enterprises of all sizes. Though extraordinarily well-funded organizations have developed sophisticated information security programs, a majority of executives still report being unsure how to confront the cybersecurity challenge.¹²⁶

The lack of publicly available information on the actual costs of previous data breaches leaves many enterprises without clear ways to evaluate the risks of a hack.¹²⁷



Simple statement of the behavior we want to change

Many executives do not invest enough time or money in addressing issues in cybersecurity, and even when they do, dollars are often misspent. Executives should prioritize cybersecurity and invest at a scale that reflects the magnitude of the issue and in ways that elicit positive bottom line security outcomes.



Behavioral Insights

>>EXECUTIVES RELY ON MENTAL MODELS THAT DO NOT ACCURATELY REFLECT THE NATURE OF CYBERSECURITY CHALLENGES. We all employ *mental models*—a shorthand description for how our minds organize abstracted knowledge, impressions, and previous experiences—that sometimes don’t translate well from one context to another. Mental models often cause us to replace one way of thinking about the world with an easier, more familiar way of thinking about the world, but too often this easier way is incomplete. Given that many executives may not have in-depth experience in information security, they’re apt to rely on mental models that lead to suboptimal decisions. For example:

- 1. ‘Good Compliance is good cybersecurity.’** Many companies have a requirement to (or feel required to) conform to externally-defined security standards (e.g. NIST). Focusing on these standards can cause executives to think that good compliance on cybersecurity standards equates to good cybersecurity. However, compliance is a necessary, not sufficient condition for security—compliance requirements can default executives to think about good cybersecurity as checking items off a list.¹²⁸ The checkbox model of security ignores the reality that standards are often outdated (technology changes faster than regulations and standards), and are often designed to achieve only a minimum threshold of security. Typically, much more needs to be accomplished than being compliant with the current minimum of standards.
- 2. ‘Cybersecurity is a technological problem; the solution must be technological.’** Part of the solution to cybersecurity is technology, but that does not mean that solving challenges in cybersecurity will require only technical solutions. A committed and well-trained team, along with clear and well-designed processes will produce better security outcomes than expensive technology investments made without complementary investments in human capital.
- 3. ‘Cybersecurity is in the weeds, and good executives do not need to get into the weeds. Therefore it’s not important to pay close attention to cybersecurity.’** There are often a lot of details related to having good cybersecurity practices, but many executives are ill-versed in the details of cybersecurity. However, just like a good executive would be able to ask pointed questions about and hold people accountable for investments in factories, marketing, or product development they need to be able to do the same when it comes to the cybersecurity of their organization. Because of risks of spear phishing and their access to sensitive data, high-level executives need to dedicate *more* time and effort to ensure their own online behaviors don’t put the enterprise at risk and set the right example for the people they are leading.

>>EXECUTIVES AVOID AMBIGUITY. This is not just true for executives. As humans, we all have *ambiguity aversion*. This is a heuristic that makes us inclined to favor options that seem to address a particular problem or problems with

well-understood risks (ones with known probabilities and outcomes). This can cause executives to both underinvest in cybersecurity and invest in the wrong items for cybersecurity. For example, consider an executive who is deciding whether to invest in a new factory or to invest in greater cybersecurity capabilities. The executive knows, within some margin of error, that investing in the factory will reduce overall costs by some factor. But, she unsure about the value of the cybersecurity investment because the risks and cost implications are highly variable. Under these circumstances, she will be much less likely to invest in cybersecurity simply because the value is ambiguous, regardless of whether it was the right thing to do. Similarly, even if an executive is making an investment in cybersecurity, they may choose to invest in products that promise to secure systems against 10,000 types of malware—a very specific number—rather than invest in redesigning processes that can provide greater improvements in overall system security in hard-to-quantify ways. This scenario would be similar to an executive at a consumer packaged-goods company choosing to fund short-term advertising campaigns that drive specific sales numbers in the short-term, but avoiding funding advertising campaigns that build general brand awareness and longer term value but that cannot be easily quantified. Unfortunately, simply telling executives that the ambiguity aversion exists does not seem to have an effect. It is a subconscious process.

>>EXECUTIVES ARE OVERCONFIDENT IN THEIR OWN ORGANIZATION'S ABILITIES TO MITIGATE CYBER RISK. In surveys, a majority of CEOs report that their organizations are doing a good job reducing cybersecurity risks, but they think their peers are doing a poor job addressing cybersecurity. Overconfidence in our own ability and outcomes is a familiar psychological phenomenon –just like all drivers consider their own driving ability to be above average (an impossibility)—this overconfidence skews an executive's appraisal of cybersecurity risk.

>>ASYMMETRIC FEEDBACK LOOP. We all respond to feedback. Feedback can come in two forms—positive feedback (we have done something well) and negative feedback (we have done something poorly). We can use this feedback to improve and make better decisions about similar situations we face in the future. When it comes to cybersecurity, the feedback loop is inconsistent and asymmetrical because attacks are rare, and when they happen, they have significant negative consequences. An executive rarely gets positive feedback

about cybersecurity because for those outside of the IT team, when a cyber attack is prevented, nothing happens—work goes on. In essence, no news is good news when it comes to cybersecurity. But, without some form of regular and positive feedback, executives may only invest in cybersecurity when receiving bad news after an attack, making it likely that they'll be able to prevent another attack of the same kind, but not a new or different kind of attack.

>>POTENTIAL FOR EXECUTIVE COMPENSATION STRUCTURE TO SKEW INVESTMENTS IN CYBERSECURITY. The structure of executive compensation levels can potentially skew executives' investment decisions when it comes to cybersecurity. Many executives' compensation gets tied to their company's performance so that if the company's stock does well, they'll often receive large bonuses (often in stock). If the company doesn't do well, the executive may earn less bonus, no bonus, or in some cases may even get fired. However, if fired, there is often a healthy severance package which pays out a significant amount of money (again, usually in stock), so much so that the executive may not need to work. But, many executives can still find another c-suite job while keeping their severance. When Boards of Directors structure CEO compensation in such a way that promotes equity (as opposed to debt) compensation it causes executives to focus their investments on upside risk opportunities (gains).¹²⁹ Cybersecurity investments by nature are focused on reducing downside risk (losses). But, because executives may get a payout regardless of whether or not the company performs well, they may focus on being risk-seeking, and neglect to make investments in cybersecurity.



Design Concepts

1 “Break” Systems. If an enterprise can afford it, consistent and constant efforts to unearth new vulnerabilities through penetration testing within internal systems gives them the best way to identify new vulnerabilities. However, this requires framing failure as a key metric for success—vulnerability detection must be encouraged, which implies a need for tolerance internally if self-attacks reveal security flaws. Focusing on breaking the system also presents the opportunity to create a positive feedback loop. For instance, by generating a report on the number of self-generated attacks stopped through this kind of testing, executives could get positive feedback for each defect found and remediated as opposed to only hearing about successful breaches from bad actors.

2 Make Security a Process, Not Just an Investment. Companies should treat compliance as a crucial first step for security, but also build in regular updates and examination of security processes to take into account the changing external threat landscape. It should be the default to have a thorough review of security processes at least annually, and one that is not just focused on compliance with existing processes and standards, but also on evaluating how current processes and standards might need to evolve to adapt to new risks.

3 Reframe Cyber Vulnerabilities as Key Business Risks.¹³⁰ Framing cybersecurity management to executives in business terms (i.e. dollar amounts and probabilities) with as much specificity as possible would facilitate a more manageable resource allocation process and reduce the abstraction of potential risks.

4 Adjust Corporate Board Focus and Role. Have the Board of Directors consider how compensation amount and structure can affect incentives related to cybersecurity, and move cybersecurity from being a discretionary decision on the part of the CEO to a mandate by the Board. This way people who are more likely to stay with the firm over the long-term (the board as opposed to the CEO) will be making the strategic decision to put cybersecurity at the top of an organization's list of priorities.

5 Survey Executives on Cybersecurity. Organizations should engage in two data collection efforts about their cybersecurity practice. First, they should regularly ask top executives outside of their company how other companies are doing on cybersecurity to provide a benchmark of performance. They should then learn about their own practices by asking those same external executives to provide a review. These two surveys could give a company's board a reasonably accurate reading on the quality of their own cybersecurity practice.

APPENDIX REFERENCES

- ¹ HP Security Research (June, 2015) HP Security Briefing: The hidden dangers of inadequate patching. Episode 22.
- ² Google. (N.D.) Android Dashboard. Accessed from <https://developer.android.com/about/dashboards/index.html> on November 14, 2016
- ³ Apple. (N.D.) App Store Support. Accessed from: <https://developer.apple.com/support/app-store/> on November 14, 2016
- ⁴ Davis, G. (June 3, 2016). Cybercriminals Crack into 360 Million Old Myspace Accounts. McAfee Securing Tomorrow Today. Accessed from: <https://securingtomorrow.mcafee.com/consumer/consumer-threat-notices/myspace-accounts-hacked/> on October 16, 2016
- ⁵ Thompson, R. F., & Spencer, W. A. (1966). Habituation: a model phenomenon for the study of neuronal substrates of behavior. *Psychological review*, 73(1), 16
- ⁶ Thaler, R. H., & Sunstein, C.R. (2008). Richard H. Thaler, Cass R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness. *Constitutional Political Economy*, 19(4), 356-360
- ⁷ ideas42 Project Team Interviews
- ⁸ HP Security Research (June, 2015) HP Security Briefing: The hidden dangers of inadequate patching. Episode 22.
- ⁹ ideas42 Project Team Interviews
- ¹⁰ Kahneman, D., Slovic, P., & Tversky, A. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185(4157), 1124-1131. <http://doi.org/10.1126/science.185.4157.1124>
- ¹¹ ideas42 Project Team Interviews
- ¹² HP Security Research (June, 2015) HP Security Briefing: The hidden dangers of inadequate patching. Episode 22.
- ¹³ <https://www.avast.com/en-us/mac>
- ¹⁴ ideas42 Project Team Interviews
- ¹⁵ ideas42 Project Team Interviews
- ¹⁶ HP Security Research (June, 2015) HP Security Briefing: The hidden dangers of inadequate patching. Episode 22.
- ¹⁷ Microsoft. (N.D.) Best Practices with Windows Server Update Services. Accessed from: [https://technet.microsoft.com/en-us/library/cc708536\(v=ws.10\).aspx](https://technet.microsoft.com/en-us/library/cc708536(v=ws.10).aspx) on November 28, 2016
- ¹⁸ iGeeks Blog (N.D.) How to enable automatic app download and update on iPhone and iPad in IOS 10. Accessed from: <https://www.igeeksblog.com/how-to-turn-on-auto-app-download-update-on-iphone-ipad/> on January 12, 2017
- ¹⁹ Apple. (Nov 7, 2016). Update the IOS software on your iPhone, iPad, or iPod Touch. Apple. Retrieved from: <https://support.apple.com/en-us/HT204204> on December 20, 2016
- ²⁰ Jenkins, J. L., Anderson, B. B., Vance, A., Kirwan, C. B., & Eargle, D. (2016). *More harm than good? How messages that interrupt can make us vulnerable*. *Information Systems Research*, 27(4), 880-896.
- ²¹ Anderson, B. B., Kirwan, C. B., Jenkins, J. L., Eargle, D., Howard, S., & Vance, A. (2015, April). How polymorphic warnings reduce habituation in the brain: Insights from an fMRI study. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2883-2892). ACM.
- ²² Modic, D., & Anderson, R. (2014). Reading this may harm your computer: The psychology of malware warnings. *Computers in Human Behavior*, 41, 71-79.
- ²³ Thor Olavsrud (Sept 2, 2014) 11 Steps Attackers Took to Crack Target. CIO. Retrieved from <http://www.cio.com/article/2600345/security/0/11-steps-attackers-took-to-crack-target.html> on October 20, 2016

- ²⁴ Anjali Rao Koppala (Nov 2, 2014) Cyberattackers Reportedly Got to JPMorgan Customer Data By Hacking The JP Morgan Corporate Challenge Website. Business Insider. Retrieved from <http://www.businessinsider.com/r-jp-morgan-found-hackers-through-breach-of-corporate-event-website-wsj-2014-10> on October 20, 2016
- ²⁵ Robert Seacord (May 5, 2014) Secure Coding to Prevent Vulnerabilities. SEI Blog. Retrieved from https://insights.sei.cmu.edu/sei_blog/2014/05/secure-coding-to-prevent-vulnerabilities.html on October 20, 2016
- ²⁶ Symantec Security Response (Sept 25, 2014) ShellShock: All you need to know about the Bash Bug vulnerability. Symantec Official Blog. Retrieved from <http://www.symantec.com/connect/blogs/shellshock-all-you-need-know-about-bash-bug-vulnerability> on October 20, 2016
- ²⁷ Hewlett-Packard (Feb, 2015) Cyber Threat Report. Hewlett-Packard Development Company L.P.
- ²⁸ RTI (May, 2002) Planning Report 02-3: The Economic Impacts of Inadequate Infrastructure for Software Testing. National Institute of Standards & Technology. Retrieved from <https://www.nist.gov/document-17633> on October 20, 2016
- ²⁹ Veracode (June, 2015) State of Software Security, Volume 6: Focus on Industry Verticals. Veracode. Retrieved from <https://info.veracode.com/state-of-software-security-report-volume6.html>
- ³⁰ Robert Lemos (April 16, 2015) App Security Worries CISOs, but Most Fail to Adopt Secure Development. Eweek. Retrieved from <http://www.eweek.com/security/app-security-worries-cisos-but-most-fail-to-adopt-secure-development.html> on October 20, 2016
- ³¹ Veracode (2016) How do vulnerabilities get into software? Veracode. Retrieved from <https://www.veracode.com/sites/default/files/Resources/Whitepapers/how-vulnerabilities-get-into-software-veracode.pdf> on October 20, 2016
- ³² Dawson, M. et al. (2010) Integrating Software Assurance into the Software Development Life Cycle (SDLC). Journal of Information Systems Technology and Planning. 3(6): 49-53
- ³³ West, D., Grant, T. (Jan 20, 2010) Agile Development: Mainstream Adoption Has Changed Agility. Forrester Research, Inc. Retrieved from: http://programmedevelopment.com/public/uploads/files/forrester_agile_development_mainstream_adoption_has_changed_agility.pdf. On October 26, 2016
- ³⁴ Queslati, H. et al. (2016) Evaluation of the Challenges of Developing Secure Software Using the Agile Approach. International Journal of Secure Software Engineering. 7(1): 17-37
- ³⁵ Veracode (2016) How do vulnerabilities get into software? Veracode. Retrieved from <https://www.veracode.com/sites/default/files/Resources/Whitepapers/how-vulnerabilities-get-into-software-veracode.pdf> on October 20, 2016
- ³⁶ Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why having too little means so much*. Macmillan.
- ³⁷ Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science*, 341(6149), 976-980.
- ³⁸ Morgan (N.D.) The Importance of Having a Secure Password [INFOGRAPHIC]. TeamsID. Access from: <https://www.teamsid.com/the-importance-of-having-a-secure-password-infographic/> on September 24, 2016
- ³⁹ ideas42 project team interviews
- ⁴⁰ RoboForm (N.D.) Password Security Survey Results – Part 1. RoboForm. Accessed from: <https://www.roboform.com/blog/password-security-survey-results> on October 2, 2016
- ⁴¹ Schneier, B. (September 5, 2014) Security of Password Managers. Schneier on Security. Accessed from: https://www.schneier.com/blog/archives/2014/09/security_of_pas.html on November 21, 2016
- ⁴² Greenberg, A. (June 15, 2015) Hack Brief: Password Manager Lastpass Got Breached Hard. Wired. Accessed from: <https://www.wired.com/2015/06/hack-brief-password-manager-lastpass-got-breached-hard/> on October 3, 2016

APPENDIX REFERENCES

⁴³ ideas42 project team interviews

⁴⁴ Hautala, L. (August 31, 2016) Dropbox Hack Leaks 68 Milion Usernames and Passwords. Cnet. Accessed from: <https://www.cnet.com/news/dropbox-hack-leaks-more-than-60-million-usernames-and-passwords/> on October 3, 2016

⁴⁵ Waqas (December 29, 2015) Top 15 Cyber Attacks and Security Breaches in 2015. HackRead. Accessed from: <https://www.hackread.com/top-15-cyber-attacks-security-breaches-in-2015/> on September 26, 2016

⁴⁶ Hamilton, A. (2015) The Top 10 Worst Security Breaches of 2014. Betanews. Accessed from: <https://betanews.com/2014/12/17/the-top-10-worst-security-breaches-of-2014/> on October 3, 2016

⁴⁷ Schneier, B. (March 9, 2009) Choosing a Bad Password Has Real-World Consequences. Schneier on Security. Accessed from: https://www.schneier.com/blog/archives/2009/03/choosing_a_bad.html on September 27, 2016

⁴⁸ ideas42 project team interviews

⁴⁹ Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of risk and uncertainty*, 1(1), 7-59.

⁵⁰ Komanduri, S., Shay, R., Kelley, P. G., Mazurek, M. L., Bauer, L., Christin, N., Conner, L. F., & Egelman, S. (2011). Of Passwords and People: Measuring the Effect of Password-Composition Policies.

⁵¹ Cowan, Nelson. "What are the differences between long-term, short-term, and working memory?." *Progress in brain research* 169 (2008): 323-338.

⁵² ideas42 project team interviews

⁵³ Glaser, A. (January 24, 2016) You Need a Password Manager. Here Are Some Good Free Ones. Wired. Accessed from: <https://www.wired.com/2016/01/you-need-a-password-manager/> on October 3, 2016

⁵⁴ Gasti, P., & Rasmussen, K. B. (2012, September). On the security of password manager database formats. In *European Symposium on Research in Computer Security* (pp. 770-787). Springer Berlin Heidelberg.

⁵⁵ Li, Z., He, W., Akhawe, D., & Song, D. (2014, July). The Emperor's New Password Manager: Security Analysis of Web-based Password Managers. In *USENIX Security* (pp. 465-479).

⁵⁶ Silver, D., Jana, S., Boneh, D., Chen, E. Y., & Jackson, C. (2014, August). Password Managers: Attacks and Defenses. In *Usenix Security* (pp. 449-464).

⁵⁷ Greenberg, A. (June 15, 2015) Hack Brief: Password Manager Lastpass Got Breached Hard. Wired. Accessed from: <https://www.wired.com/2015/06/hack-brief-password-manager-lastpass-got-breached-hard/> on October 3, 2016

⁵⁸ Schneier, B. (N.D.) Password Safe: A security of Twofish in a Password Database. Schneier on Security. Accessed from: <https://www.schneier.com/academic/passsafe/> on December 21, 2016

⁵⁹ Schneier, B. (September 5, 2014) Security of Password Managers. Schneier on Security. Accessed from: https://www.schneier.com/blog/archives/2014/09/security_of_pas.html on November 21, 2016

⁶⁰ Komanduri, S., Shay, R., Kelley, P. G., Mazurek, M. L., Bauer, L., Christin, N., Conner, L. F., & Egelman, S. (2011). Of Passwords and People: Measuring the Effect of Password-Composition Policies.

⁶¹ Komanduri, S., Shay, R., Kelley, P. G., Mazurek, M. L., Bauer, L., Christin, N., Conner, L. F., & Egelman, S. (2011). Of Passwords and People: Measuring the Effect of Password-Composition Policies.

⁶² This concept is helpfully illustrated by the popular comic, XKCD: <https://xkcd.com/936/>

⁶³ Schneier, B. (March 3, 2014) Choosing Secure Passwords. Schneier on Security. Accessed from: https://www.schneier.com/blog/archives/2014/03/choosing_secure_1.html on October 26, 2016

⁶⁴ Schneier, B. (March 3, 2014) Choosing Secure Passwords. Schneier on Security. Accessed from: https://www.schneier.com/blog/archives/2014/03/choosing_secure_1.html on October 26, 2016

- ⁶⁵ Thompson, R. F., & Spencer, W. A. (1966). Habituation: a model phenomenon for the study of neuronal substrates of behavior. *Psychological review*, 73(1), 16.
- ⁶⁶ Wisner, M. (October 18, 2016) Netscape Co-Founder: Passwords Are the Weak Link in Cyber Security. Fox Business. Accessed from: <https://www.yahoo.com/tech/netscape-co-founder-passwords-weak-150300764.html> on October 27, 2016
- ⁶⁷ LaunchKey (August 12, 2015) Password Survey – Preliminary Results. LaunchKey. Accessed from: https://s3.amazonaws.com/launchkey-blog/LaunchKey_Password_Survey_Results.pdf on October 3, 2016
- ⁶⁸ Yegulalp, S. (September 25, 2014) Dump Your Passwords! 8 Security and Identity Breakthroughs. InfoWorld. Accessed from: <http://www.infoworld.com/article/2687352/password-security/password-security-164550-8-cutting-edge-technologies-aimed-at-eliminating-passwords.html#slide1> on October 25, 2016
- ⁶⁹ Hern, A. (May 24, 2016) Google Aims to Kill Passwords by the End of the Year. The Guardian. Accessed from: <https://www.theguardian.com/technology/2016/may/24/google-passwords-android> on September 25, 2016
- ⁷⁰ Petsas, T., Tsirantonakis, G., Athanasopoulos, E., & Ioannidis, S. (2015, April). Two-factor authentication: is the world ready?: quantifying 2FA adoption. In *Proceedings of the Eighth European Workshop on System Security* (p. 4). ACM.
- ⁷¹ Krebs, B. (September, 2016) The Limits of SMS for 2-Factor Authentication. Krebs on Security. Accessed from: <https://krebsonsecurity.com/2016/09/the-limits-of-sms-for-2-factor-authentication/> on October 26, 2016
- ⁷² Ross, R., White, S., Wright, J., & Knapp, L. (2013, May). Using behavioral economics for postsecondary success. In *Ideas* (Vol. 42).
- ⁷³ Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The journal of economic perspectives*, 5(1), 193-206.
- ⁷⁴ Symantec. (2016). *Internet Security Threat Report*.
- ⁷⁵ Cisco. (2008). *Data leakage worldwide: the effectiveness of security policies*.
- ⁷⁶ Herath, T., & Rao, H. R. (2009). Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems*, 47(2), 154-165.
- ⁷⁷ Stanton, J. M., Stam, K. R., Mastrangelo, P., & Jolton, J. (2005). Analysis of end user security behaviors. *Computers & Security*, 24(2), 124-133.
- ⁷⁸ Cisco. (2008). *Data leakage worldwide: the effectiveness of security policies*.
- ⁷⁹ Meyer, J. P., & Allen, N. J. (1991). A three-component conceptualization of organizational commitment. *Human resource management review*, 1(1), 61-89.
- ⁸⁰ Cialdini, R. B. (1987). *Influence* (Vol. 3). A. Michel.
- ⁸¹ Das, S., Kim, T. H. J., Dabbish, L. A., & Hong, J. I. (2014, July). The effect of social influence on security sensitivity. In *Proc. SOUPS* (Vol. 14).
- ⁸² Herath, T., & Rao, H. R. (2009). Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems*, 47(2), 154-165.
- ⁸³ Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration.
- ⁸⁴ Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of personality and social psychology*, 8(4), 377-383.
- ⁸⁵ Herath, T., & Rao, H. R. (2009). Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems*, 47(2), 154-165.
- ⁸⁶ Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2010). Getting to the top of mind: How reminders increase saving (No. w16205). National Bureau of Economic Research.
- ⁸⁷ Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why having too little means so much*. Macmillan.

APPENDIX REFERENCES

- ⁸⁸ Meyer, J. P., & Allen, N. J. (1991). A three-component conceptualization of organizational commitment. *Human resource management review*, 1(1), 61-89.
- ⁸⁹ Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2010). Getting to the top of mind: How reminders increase saving (No. w16205). National Bureau of Economic Research.
- ⁹⁰ For a longer list of the elements of access controls, see: <http://www.nationalcybersecurityinstitute.org/general-public-interests/cybersecurity-access-control/>
- ⁹¹ Bailey, B. (2015, February 6). *Anthem: Hackers Tried to Breach System as Early as Dec. 10*. Retrieved from <http://www.usnews.com/news/business/articles/2015/02/06/anthem-hacker-tried-to-breach-system-as-early-as-dec-10>
- ⁹² Radichel, T. (2014). Case Study: Critical Controls that Could Have Prevented Target Breach. *SANS Institute InfoSec Reading Room*.
- ⁹³ Weingarten, E. (2016, May 16). *Burn, Baby Burn*. Retrieved from <https://context.newamerica.org/burn-baby-burn-d99f08b293be#.p518gqx2r>
- ⁹⁴ Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why having too little means so much*. Macmillan.
- ⁹⁵ Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly journal of experimental psychology*, 12(3), 129-140.
- ⁹⁶ Verizon. (2016) *Data Breach Investigations Report*
- ⁹⁷ United Nations Office on Drugs and Crime (2013) *Comprehensive Study on Cybercrime*
- ⁹⁸ United Nations Office on Drugs and Crime (2013) *Comprehensive Study on Cybercrime*
- ⁹⁹ Exec. Order No. 13636, 3 C.F.R., 2013
- ¹⁰⁰ Choucri, N., Madnick, S., & Koepke, P. (2016). Institutions for Cyber Security: International Responses and Data Sharing Initiatives.
- ¹⁰¹ Intel Security. (2016). *McAfee Labs Threats Report*
- ¹⁰² Choucri, N., Madnick, S., & Koepke, P. (2016). Institutions for Cyber Security: International Responses and Data Sharing Initiatives.
- ¹⁰³ Intel Security. (2016). *McAfee Labs Threats Report*
- ¹⁰⁴ Hulme, G. B. (2017, January 17) *Tackling cybersecurity threat information sharing challenges*. Retrieved from <http://www.csoonline.com/article/3157540/security/tackling-cybersecurity-threat-information-sharing-challenges.html>
- ¹⁰⁵ Petrasic, K. & Bornfreund, M., (2016, April 18) *CISA Guidance Clarifies How to Share Cyber Threat Information... but Issues Remain*. Retrieved from <http://www.whitecase.com/publications/alert/cisa-guidance-clarifies-how-share-cyber-threat-information-issues-remain>
- ¹⁰⁶ Kvochko, E. & Pant, R. (2017). Why Data Breaches Don't Hurt Stock Prices." *Harvard Business Review*.
- ¹⁰⁷ Madejski, M., Johnson, M. and Bellovin, S. A Study of Privacy Settings Errors in an Online Social Network. Retrieved from <https://www.cs.columbia.edu/~smb/papers/fb-violations-sesoc.pdf>
- ¹⁰⁸ Kirlappos, I., Sasse, M. A. (2012). Security Education against Phishing: A Modest Proposal for a Major Rethink. *IEEE Security and Privacy Magazine* 10(2), 24-32
- ¹⁰⁹ Iyengar, S.S.; Lepper, M.R. (2000). "When choice is demotivating: can one desire too much of a good thing?". *Journal of Personality and Social Psychology*. 79: 995–1006
- ¹¹⁰ Schwartz, B. (2004, January). *The paradox of choice: Why more is less*. New York: Ecco.
- ¹¹¹ Debatin, B., Lovejoy, J., Horn, A., Hughes, B. (2009). Facebook and Online Privacy: Attitudes, Behaviors and Unintended Consequences. *Journal of Computer-Mediated Communication*. 15(1), 83-108.

- ¹¹² Das, S., Kramer, A. D., Dabbish, L. A., & Hong, J. I. (2014, November). Increasing security sensitivity with social proof: A large-scale experimental confirmation. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security* (pp. 739-749). ACM.
- ¹¹³ APWG (Feb, 2017). Phishing Activity Trends Report 4th QR 2016. APWG. Accessed from: http://www.antiphishing.org/resources/apwg-reports/apwg_trends_report_q4_2016.pdf on March 3, 2017
- ¹¹⁴ Benenson, Z., Gassmann, F., & Landwirth, R. (2016) Exploiting curiosity and context: How to make people click on a dangerous link despite their security awareness. Accessed from: <http://paper.seebug.org/papers/Security%20Conf/Blackhat/2016/us-16-Benenson-Exploiting-Curiosity-And-Context-How-To-Make-People-Click-On-A-Dangerous-Link-Despite-Their-Security-Awareness-wp.pdf>
- ¹¹⁵ Cloudmark (2016) Survey Revelas Spear Phishing as Top Security Concern to Enterprises. *Cloudmark Security Blog*. Accessed from: <https://blog.cloudmark.com/2016/01/13/survey-spear-phishing-a-top-security-concern-to-enterprises/> on October 16, 2016
- ¹¹⁶ Mandia, K. Testimony before the U.S. Senate Select Committee on Intelligence. 30 March 2017. Available at: <https://www.intelligence.senate.gov/hearings/open-hearing-intelligence-matters-1>. Accessed 3/30/2017.
- ¹¹⁷ Milgram, S. (1963). Behavioral Study of obedience. *The Journal of abnormal and social psychology*, 67(4), 371.
- ¹¹⁸ Cialdini, R. B. (2007). *Influence: The psychology of persuasion*. New York: Collins.
- ¹¹⁹ Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of personality and social psychology*, 9(2p2), 1.
- ¹²⁰ Vishwanath, A., Herath, T., Chen, R., Wang, J., & Rao, H. R. (2011). Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems*, 51(3), 576-586.
- ¹²¹ Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The journal of economic perspectives*, 5(1), 193-206.
- ¹²² Dhamija, R., Tygar, J. D., & Hearst, M. (2006, April). Why phishing works. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 581-590). ACM.
- ¹²³ Benenson, Z. Exploiting Curiosity and Context: How to make people click on a dangerous link despite their security awareness. Retrieved from <https://www.blackhat.com/docs/us-16/materials/us-16-Benenson-Exploiting-Curiosity-And-Context-How-To-Make-People-Click-On-A-Dangerous-Link-Despite-Their-Security-Awareness.pdf>
- ¹²⁴ Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2010). Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology (TOIT)*, 10(2), 7.
- ¹²⁵ Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A., & Pham, T. (2009, July). School of phish: a real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (p. 3). ACM.
- ¹²⁶ Marshall, M., Bell, G., Michaux, D., Bernd-Striebeck, U., Archibald, G., Glines, S. (2015). Cyber security: a failure of imagination by CEOs Global CEOs walk a fine line between risk and reward, 1-12.
- ¹²⁷ Gordon, L. A., & Loeb, M. P. (2002). The economics of information security investment. *ACM Transactions on Information and System Security (TISSEC)*, 5(4), 438-457.
- ¹²⁸ Moore, T., Dynes, S., & Chang, F. R. (2015). Identifying how firms manage cybersecurity investment. Available: Southern Methodist University. Available at: <http://blog.smu.edu/research/files/2015/10/SMU-IBM.pdf> (Accessed 2015-12-14), 32.
- ¹²⁹ Eisdorfer, A., Giaccotto, C., & White, R. (2013). Capital structure, executive compensation, and investment efficiency. *Journal of Banking & Finance*, 37(2), 549-562.
- ¹³⁰ ideas42 Project Team Interview

When a rogue press release defaming the CEO of a major corporation is published, apparently by the company's own communications department, the race is on to figure out what happened—and how. As the FBI scrutinizes whether big investments in secure technology deliver the iron fortress around sensitive data that they promise, a closer look at the movements and perspective of the hacker—who is as skilled at predicting human behavior as she is at getting past firewalls—may shed light on the true fragilities in the system.

This novella was produced by ideas42 with generous support from the William and Flora Hewlett Foundation in an effort to highlight the human factor that is often left out of stories of cybersecurity breaches. It also includes insights into some unexpected barriers when it comes to keeping online data private and secure on individual and organizational levels.



ABOUT IDEAS42

At ideas42 we believe that a deep understanding of human behavior will help us improve millions of lives. Using insights from behavioral science, we create innovative solutions in economic mobility, health, education, criminal justice, consumer finance, energy efficiency and international development. We're a nonprofit with more than 80 active projects in the United States and around the world and many partnerships across governments, foundations, NGOs and corporations.



To find out more, visit us at ideas42.org
or follow us @ideas42  